



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<p>(51) International Patent Classification ⁶ : C12N 15/31, C07K 14/315, A61K 39/09, C12N 1/21</p>	<p>A2</p>	<p>(11) International Publication Number: WO 99/42588</p> <p>(43) International Publication Date: 26 August 1999 (26.08.99)</p>
<p>(21) International Application Number: PCT/CA99/00114</p> <p>(22) International Filing Date: 17 February 1999 (17.02.99)</p> <p>(30) Priority Data: 60/075,425 20 February 1998 (20.02.98) US</p> <p>(71) Applicant (for all designated States except US): BIOCHEM VACCINS INC. [CA/CA]; 2323 boulevard du Parc Technologique, Sainte-Foy, Québec G1P 4R8 (CA).</p> <p>(72) Inventors; and (75) Inventors/Applicants (for US only): BRODEUR, Bernard, R. [CA/CA]; 2401 rue Maritain, Sillery, Québec G1T 1N6 (CA). RIOUX, Clément [CA/CA]; 1012 Jean-Charles Cantin, Ville de Cap Rouge, Québec G1Y 2X1 (CA). BOYER, Martine [CA/CA]; Apt. 204, 25 des Mouettes, Beauport, Québec G1E 7G1 (CA). CHARLEBOIS, Isabelle [CA/CA]; 410 Mirabel, St-Nicolas, Québec G7A 2L5 (CA). HAMEL, Josée [CA/CA]; 2401 rue Maritain, Sillery, Québec G1T 1N6 (CA). MARTIN, Denis [CA/CA]; 4728-G rue Gaboury, St-Augustin-de-Desmaures, Québec G3A 1E9 (CA).</p>	<p>(74) Agents: CÔTE, France et al.; Swabey Ogilvy Renault, Suite 1600, 1981 McGill College Avenue, Montréal, Québec H3A 2Y3 (CA).</p> <p>(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).</p> <p>Published <i>Without international search report and to be republished upon receipt of that report.</i></p>	
<p>(54) Title: GROUP B STREPTOCOCCUS ANTIGENS</p>		
<p>(57) Abstract</p> <p>Group B streptococcus (GBS) proteins and polynucleotides encoding them are disclosed. Said proteins are antigenic and therefore useful vaccine components for the prophylaxis or therapy of streptococcus infection in animals. Also disclosed are recombinant methods of producing the protein antigens as well as diagnostic assays for detecting streptococcus bacterial infection.</p>		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

GROUP B STREPTOCOCCUS ANTIGENS

5

FIELD OF THE INVENTION

The present invention is related to antigens, more particularly protein antigens of group B streptococcus (GBS) bacterial pathogen which are useful as vaccine components for therapy and/or prophylaxis.

10

BACKGROUND OF THE INVENTION

15

Streptococcus are gram (+) bacteria that are differentiated by group specific carbohydrate antigens A through O found on their cell surface. Streptococcus groups are further distinguished by type-specific capsular polysaccharide antigens. Several serotypes have been identified for the Group B streptococcus (GBS) : Ia, Ib, II, III, IV, V, VI, VII and VIII. GBS also contains antigenic proteins known as "C-proteins" (alpha, beta, gamma and delta), some of which have been cloned.

25

Although GBS is a common component of the normal human vaginal and colonic flora this pathogen has long been recognized as a major cause of neonatal sepsis and meningitis, late-onset meningitis in infants, postpartum endometritis as well as mastitis in dairy herds. Expectant mothers exposed to GBS are at risk of postpartum infection and may transfer the infection to their baby as the child passes through the birth canal. Although the organism is sensitive to antibiotics, the high attack rate and rapid onset of sepsis in neonates and meningitis in infants results in high morbidity and mortality.

35

To find a vaccine that will protect individuals from GBS infection, researches have turned to the type-specific antigens. Unfortunately these polysaccharides have proven to
5 be poorly immunogenic in humans and are restricted to the particular serotype from which the polysaccharide originates. Further, capsular polysaccharide elicit a T cell independent response i.e. no IgG production. Consequently capsular polysaccharide antigens are unsuitable
10 as a vaccine component for protection against GBS infection.

Others have focused on the C-protein beta antigen which demonstrated immunogenic properties in mice and rabbit models. This protein was found to be unsuitable as a human
15 vaccine because of its undesirable property of interacting with high affinity and in a non-immunogenic manner with the Fc region of human IgA. The C-protein alpha antigen is rare in type III serotypes of GBS which is the serotype responsible for most GBS mediated conditions and is
20 therefore of little use as a vaccine component.

Therefore there remains an unmet need for GBS antigens that may be used as vaccine components for the prophylaxis and/or
25 therapy of GBS infection.

SUMMARY OF THE INVENTION

30 According to one aspect, the present invention provides an isolated polynucleotide encoding a polypeptide having at least 70% identity to a second polypeptide comprising a sequence selected from the group consisting of:

35 SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5,
SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO:10,
SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:15,
SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19,

SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:24,
SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:29,
SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:34,
SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:39,
5 SEQ ID NO:40, SEQ ID NO:41 and SEQ ID NO:44 or fragments,
analogous or derivatives thereof.

In other aspects, there is provided vectors comprising
polynucleotides of the invention operably linked to an
10 expression control region, as well as host cells transfected
with said vectors and methods of producing polypeptides
comprising culturing said host cells under conditions
suitable for expression.

15 In yet another aspect, there is provided novel polypeptides
encoded by polynucleotides of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

20

Figure 1a is the DNA sequence of clone 1 (SEQ ID NO :1) with
corresponding amino acid sequences for open reading frames;
figure 1b is the amino acid sequence SEQ ID NO: 2;
figure 1c is the amino acid sequence SEQ ID NO: 3;
25 figure 1d is the amino acid sequence SEQ ID NO: 4;
figure 1e is the amino acid sequence SEQ ID NO: 5;
figure 1f is the amino acid sequence SEQ ID NO: 6;

Figure 2a is the DNA sequence of clone 2 (SEQ ID NO :7) with
30 corresponding amino acid sequences for open reading frames;
figure 2b is the amino acid sequence SEQ ID NO: 8;
figure 2c is the amino acid sequence SEQ ID NO: 9;
figure 2d is the amino acid sequence SEQ ID NO:10;
figure 2e is the amino acid sequence SEQ ID NO:11;
35 figure 2f is the amino acid sequence SEQ ID NO:12;

Figure 3a is the DNA sequence of clone 3 (SEQ ID NO :13) with corresponding amino acid sequences for open reading frames;

figure 3b is the amino acid sequence SEQ ID NO:14;

5 figure 3c is the amino acid sequence SEQ ID NO:15;

figure 3d is the amino acid sequence SEQ ID NO:16;

figure 3e is the amino acid sequence SEQ ID NO:17;

figure 3f is the amino acid sequence SEQ ID NO:18;

figure 3g is the amino acid sequence SEQ ID NO:19;

10 figure 3h is the amino acid sequence SEQ ID NO:20;

figure 3i is the amino acid sequence SEQ ID NO:21;

Figure 4a is the DNA sequence of clone 4 (SEQ ID NO :22) with corresponding amino acid sequences for open reading frames;

15 figure 4b is the amino acid sequence SEQ ID NO:23;

figure 4c is the amino acid sequence SEQ ID NO:24;

figure 4d is the amino acid sequence SEQ ID NO:25;

20 figure 4e is the amino acid sequence SEQ ID NO:26;

Figure 5a is the DNA sequence of clone 5 (SEQ ID NO :27) with corresponding amino acid sequences for open reading frames;

figure 5b is the amino acid sequence SEQ ID NO:28;

25 figure 5c is the amino acid sequence SEQ ID NO:29;

figure 5d is the amino acid sequence SEQ ID NO:30;

figure 5e is the amino acid sequence SEQ ID NO:31;

Figure 6a is the DNA sequence of clone 6 (SEQ ID NO :32) ;

30 figure 6b is the amino acid sequence SEQ ID NO:33;

figure 6c is the amino acid sequence SEQ ID NO:34;

figure 6d is the amino acid sequence SEQ ID NO:35;

figure 6e is the amino acid sequence SEQ ID NO:36;

35 Figure 7a is the DNA sequence of clone 7 (SEQ ID NO :37);

figure 7b is the amino acid sequence SEQ ID NO:38;

figure 7c is the amino acid sequence SEQ ID NO:39;

figure 7d is the amino acid sequence SEQ ID NO:40;

figure 7e is the amino acid sequence SEQ ID NO:41;

- 5 Figure 8 is the DNA sequence of a part of clone 7 including a signal sequence (SEQ ID NO :42);

Figure 9 is the DNA sequence of a part of clone 7 without a signal sequence (SEQ ID NO :43);

- 10 Figure 9a is the amino acid sequence (SEQ ID NO:44);

Figure 10 represents the distribution of anti-GBS ELISA titers in sera from CD-1 mice immunized with recombinant GBS protein corresponding to the SEQ ID NO:39.

DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to novel antigenic polypeptides of group B streptococcus (GBS) characterized by the amino acid sequence selected from the group consisting of:

5 SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5,
SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO:10,
SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:15,
10 SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19,
SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:24,
SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:29,
SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:34,
SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:39,
15 SEQ ID NO:40, SEQ ID NO:41 and SEQ ID NO:44 or fragments,
analogues or derivatives thereof.

A preferred embodiment of the invention includes SEQ ID NO :39 and SEQ ID NO:44.

20 A further preferred embodiment of the invention is SEQ ID NO :39.

A further preferred embodiment of the invention is SEQ ID NO :44.

25 As used herein, "fragments", "derivatives" or "analogues" of the polypeptides of the invention include those polypeptides in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably conserved) and which may be natural or unnatural.

30 The terms «fragments», «derivatives» or «analogues» of polypeptides of the present invention also include polypeptides which are modified by addition, deletion,

substitution of amino acids provided that the polypeptides retain the capacity to induce an immune response.

5 By the term «conserved amino acid» is meant a substitution of one or more amino acids for another in which the antigenic determinant (including its secondary structure and hydrophobic nature) of a given antigen is completely or partially conserved in spite of the substitution.

10 For example, one or more amino acid residues within the sequence can be substituted by another amino acid of a similar polarity, which acts as a functional equivalent, resulting in a silent alteration. Substitutes for an amino acid within the sequence may be selected from other members
15 of the class to which the amino acid belongs. For example, the nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan and methionine. The polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine,
20 asparagine and glutamine. The positively charged (basic) amino acids include arginine, lysine and histidine. The negatively charged (acidic) amino acids include aspartic acid and glutamic acid.

25 Preferably, derivatives and analogs of polypeptides of the invention will have about 70% identity with those sequences illustrated in the figures or fragments thereof. That is, 70% of the residues are the same. More preferably polypeptides will have greater than 95% homology. In another
30 preferred embodiment, derivatives and analogs of polypeptides of the invention will have fewer than about 20 amino acid residue substitutions, modifications or deletions and more preferably less than 10. Preferred substitutions are those known in the art as conserved i.e. the substituted
35 residues share physical or chemical properties such as hydrophobicity, size, charge or functional groups.

Furthermore, in those situations where amino acid regions are found to be polymorphic, it may be desirable to vary one or more particular amino acids to more effectively mimic the
5 different epitopes of the different GBS strains.

Also included are polypeptides which have fused thereto other compounds which alter the polypeptides biological or pharmacological properties i.e. polyethylene glycol (PEG) to
10 increase half-life; leader or secretory amino acid sequences for ease of purification; prepro- and pro- sequences; and (poly)saccharides.

Moreover, the polypeptides of the present invention can be
15 modified by terminal $-NH_2$ acylation (eg. by acetylation, or thioglycolic acid amidation, terminal carboxy amidation, e.g. with ammonia or methylamine) to provide stability, increased hydrophobicity for linking or binding to a support or other molecule.

20 Also contemplated are hetero and homo polypeptide multimers of the polypeptide fragments, analogues and derivatives. These polymeric forms include, for example, one or more polypeptides that have been cross-linked with cross-linkers
25 such as avidin/biotin, gluteraldehyde or dimethyl-superimidate. Such polymeric forms also include polypeptides containing two or more tandem or inverted contiguous sequences, produced from multicistronic mRNAs generated by recombinant DNA technology.

30 Preferably, a fragment, analog or derivative of a polypeptide of the invention will comprise at least one antigenic region i.e. at least one epitope.

In order to achieve the formation of antigenic polymers
35 (i.e. synthetic multimers), polypeptides may be utilized having bishaloacetyl groups, nitroarylhalides, or the like,

where the reagents being specific for thio groups.
Therefore, the link between two mercapto groups of the
different peptides may be a single bond or may be composed
of a linking group of at least two, typically at least four,
5 and not more than 16, but usually not more than about 14
carbon atoms.

In a particular embodiment, polypeptide fragments, analogs
and derivatives of the invention do not contain a methionine
10 (Met) starting residue. Preferably, polypeptides will not
incorporate a leader or secretory sequence (signal
sequence). The signal portion of a polypeptide of the
invention may be determined according to established
molecular biological techniques. In general, the
15 polypeptide of interest may be isolated from a GBS culture
and subsequently sequenced to determine the initial residue
of the mature protein and therefor the sequence of the
mature polypeptide.

20 According to another aspect, there is provided vaccine
compositions comprising one or more GBS polypeptides of the
invention in admixture with a pharmaceutically acceptable
carrier diluent or adjuvant.

25 Suitable adjuvants include oils i.e. Freund's complete or
incomplete adjuvant; salts i.e. $AlK(SO_4)_2$, $AlNa(SO_4)_2$,
 $AlNH_4(SO_4)_2$, $Al(OH)_3$, $AlPO_4$, silica, kaolin; saponin
derivative; carbon polynucleotides i.e. poly IC and poly AU
and also detoxified cholera toxin (CTB) and E.coli heat
30 labile toxin for induction of mucosal immunity. Preferred
adjuvants include QuilA™, Alhydrogel™ and Adjuphos™.
Vaccines of the invention may be administered parenterally
by injection, rapid infusion, nasopharyngeal absorption,
dermoabsorption, or bucal or oral.

35

Vaccine compositions of the invention are used for the treatment or prophylaxis of *streptococcus* infection and/or diseases and symptoms mediated by *streptococcus* infection, in particular group A *streptococcus* (*pyogenes*), group B *streptococcus* (GBS or *agalactiae*), *dysgalactiae*, *uberis*, *nocardia* as well as *Staphylococcus aureus*. General information about *Streptococcus* is available in Manual of Clinical Microbiology by P.R.Murray et al. (1995, 6th Edition, ASM Press, Washington, D.C.). More particularly group B *streptococcus*, *agalactiae*. In a particular embodiment vaccines are administered to those individuals at risk of GBS infection such as pregnant women and infants for sepsis, meningitis and pneumonia as well as immunocompromised individuals such as those with diabetes, liver disease or cancer. Vaccines may also have veterinary applications such as for the treatment of mastitis in cattle which is mediated by the above mentioned bacteria as well as *E.coli*.

The vaccine of the present invention can also be used for the manufacture of a medicament used for the treatment or prophylaxis of *streptococcus* infection and/or diseases and symptoms mediated by *streptococcus* infection, in particular group A *streptococcus* (*pyogenes*), group B *streptococcus* (GBS or *agalactiae*), *dysgalactiae*, *uberis*, *nocardia* as well as *Staphylococcus aureus*. More particularly group B *streptococcus*, *agalactiae*.

Vaccine compositions are preferably in unit dosage form of about 0.001 to 100 µg/kg (antigen/body weight) and more preferably 0.01 to 10 µg/kg and most preferably 0.1 to 1 µg/kg 1 to 3 times with an interval of about 1 to 12 weeks intervals between immunizations, and more preferably 1 to 6

weeks.

According to another aspect, there is provided polynucleotides encoding polypeptides of group B

5 streptococcus (GBS) characterized by the amino acid sequence selected from the group consisting of:

SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5,
SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO:10,
SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:15,
10 SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19,
SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:24,
SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:29,
SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:34,
SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:39,
15 SEQ ID NO:40, SEQ ID NO:41 and SEQ ID NO:44 or fragments,
analogs or derivatives thereof.

Preferred polynucleotides are those illustrated in figures
1a (SEQ ID NO: 1), 2a (SEQ ID NO: 7), 3a (SEQ ID NO: 13), 4a
20 (SEQ ID NO: 22), 5a (SEQ ID NO: 27), 6a (SEQ ID NO: 32), 7a
(SEQ ID NO: 37), 8 (SEQ ID NO : 42) and 9 (SEQ ID NO : 43)
which correspond to the open reading frames, encoding
polypeptides of the invention.

25 Preferred polynucleotides are those illustrated in figures
1a (SEQ ID NO: 1), 2a (SEQ ID NO: 7), 3a (SEQ ID NO: 13), 4a
(SEQ ID NO: 22), 5a (SEQ ID NO: 27), 6a (SEQ ID NO: 32), 7a
(SEQ ID NO: 37), 8 (SEQ ID NO : 42) and 9 (SEQ ID NO : 43)
and fragments, analogues and derivatives thereof.

30 More preferred polynucleotides of the invention are those
illustrated in Figures 7 (SEQ ID NO : 37), 8 (SEQ ID NO :
42) and 9 (SEQ ID NO : 43).

35 Most preferred polynucleotides of the invention are those
illustrated in Figures 8 (SEQ ID NO : 42) and 9 (SEQ ID NO :

43) .

It will be appreciated that the polynucleotide sequences illustrated in the figures may be altered with degenerate codons yet still encode the polypeptides of the invention.

Due to the degeneracy of nucleotide coding sequences, other polynucleotide sequences which encode for substantially the same polypeptides of the present invention may be used in the practice of the present invention. These include but are not limited to nucleotide sequences which are altered by the substitution of different codons that encode the same amino acid residue within the sequence, thus producing a silent change.

Accordingly the present invention further provides polynucleotides which hybridize to the polynucleotide sequences herein above described (or the complement sequences thereof) having 50% and preferably at least 70% identity between sequences. More preferably polynucleotides are hybridizable under stringent conditions i.e. having at least 95% identity and most preferably more than 97% identity.

By capable of hybridizing under stringent conditions is meant annealing of a nucleic acid molecule to at least a region of a second nucleic acid sequence (whether as cDNA, mRNA, or genomic DNA) or to its complementary strand under standard conditions, e.g. high temperature and/or low salt content, which tend to disfavor hybridization of noncomplementary nucleotide sequences. A suitable protocol is described in Maniatis T. et al., Molecular cloning : A Laboratory Manual, Cold Springs Harbor Laboratory, 1982, which is herein incorporated by reference.

In a further aspect, polynucleotides encoding polypeptides

of the invention, or fragments, analogs or derivatives thereof, may be used in a DNA immunization method. That is, they can be incorporated into a vector which is replicable and expressible upon injection thereby producing
5 the antigenic polypeptide in vivo. For example polynucleotides may be incorporated into a plasmid vector under the control of the CMV promoter which is functional in eukaryotic cells. Preferably the vector is injected intramuscularly.

10

According to another aspect, there is provided a process for producing polypeptides of the invention by recombinant techniques by expressing a polynucleotide encoding said polypeptide in a host cell and recovering the expressed
15 polypeptide product. Alternatively, the polypeptides can be produced according to established synthetic chemical techniques i.e. solution phase or solid phase synthesis of oligopeptides which are ligated to produce the full polypeptide (block ligation).

20

For recombinant production, host cells are transfected with vectors which encode the polypeptide, and then cultured in a nutrient media modified as appropriate for activating promoters, selecting transformants or amplifying the genes.
25 Suitable vectors are those that are viable and replicable in the chosen host and include chromosomal, non-chromosomal and synthetic DNA sequences e.g. bacterial plasmids, phage DNA, baculovirus, yeast plasmids, vectors derived from combinations of plasmids and phage DNA. The polypeptide
30 sequence may be incorporated in the vector at the appropriate site using restriction enzymes such that it is operably linked to an expression control region comprising a promoter, ribosome binding site (consensus region or Shine-Dalgarno sequence), and optionally an operator (control
35 element). One can select individual components of the expression control region that are appropriate for a given

host and vector according to established molecular biology principles (Sambrook et al, Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor, N.Y., 1989 incorporated herein by reference). Suitable promoters include but are not limited to LTR or SV40 promoter, *E.coli* lac, tac or trp promoters and the phage lambda P_L promoter. Vectors will preferably incorporate an origin of replication as well as selection markers i.e. ampicillin resistance gene. Suitable bacterial vectors include pET, pQE70, pQE60, pQE-9, pbs, pD10 phagescript, psiX174, pbluescript SK, pbsks, pNH8A, pNH16a, pNH18A, pNH46A, ptrc99a, pKK223-3, pDR540, pPRIT5 and eukaryotic vectors pBlueBacIII, pWLNEO, pSV2CAT, pOG44, pXT1, pSG, pSVK3, pBPV, pMSG and pSVL. Host cells may be bacterial i.e. *E.coli*, *Bacillus subtilis*, *Streptomyces*; fungal i.e. *Aspergillus niger*, *Aspergillus nidulins*; yeast i.e. *Saccharomyces* or eukaryotic i.e. CHO, COS.

Upon expression of the polypeptide in culture, cells are typically harvested by centrifugation then disrupted by physical or chemical means (if the expressed polypeptide is not secreted into the media) and the resulting crude extract retained to isolate the polypeptide of interest. Purification of the polypeptide from culture media or lysate may be achieved by established techniques depending on the properties of the polypeptide i.e. using ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, hydroxylapatite chromatography and lectin chromatography. Final purification may be achieved using HPLC.

The polypeptide may be expressed with or without a leader or secretion sequence. In the former case the leader may be removed using post-translational processing (see US

4,431,739; 4,425,437; and 4,338,397 incorporated herein by reference) or be chemically removed subsequent to purifying the expressed polypeptide.

- 5 According to a further aspect, the GBS polypeptides of the invention may be used in a diagnostic test for streptococcus infection in particular GBS infection. Several diagnostic methods are possible, for example detecting streptococcus organism in a biological sample, the following procedure may
- 10 be followed:
- a) obtaining a biological sample from a patient;
 - b) incubating an antibody or fragment thereof reactive with a GBS polypeptide of the invention with the biological sample to form a mixture; and
 - 15 c) detecting specifically bound antibody or bound fragment in the mixture which indicates the presence of streptococcus.

Alternatively, a method for the detection of antibody

20 specific to a streptococcus antigen in a biological sample containing or suspected of containing said antibody may be performed as follows:

- a) isolating a biological sample from a patient;
- b) incubating one or more GBS polypeptides of the
- 25 invention or fragments thereof with the biological sample to form a mixture; and
- c) detecting specifically bound antigen or bound fragment in the mixture which indicates the presence of antibody specific to streptococcus.

30 One of skill in the art will recognize that this diagnostic test may take several forms, including an immunological test such as an enzyme-linked immunosorbent assay (ELISA), a radioimmunoassay or a latex agglutination assay, essentially

35 to determine whether antibodies specific for the protein are present in an organism.

The DNA sequences encoding polypeptides of the invention may also be used to design DNA probes for use in detecting the presence of streptococcus in a biological sample suspected of containing such bacteria. The detection method of this invention comprises:

- a) isolating the biological sample from a patient;
- b) incubating one or more DNA probes having a DNA sequence encoding a polypeptide of the invention or fragments thereof with the biological sample to form a mixture; and
- c) detecting specifically bound DNA probe in the mixture which indicates the presence of streptococcus bacteria.

The DNA probes of this invention may also be used for detecting circulating streptococcus i.e. GBS nucleic acids in a sample, for example using a polymerase chain reaction, as a method of diagnosing streptococcus infections. The probe may be synthesized using conventional techniques and may be immobilized on a solid phase, or may be labeled with a detectable label. A preferred DNA probe for this application is an oligomer having a sequence complementary to at least about 6 contiguous nucleotides of the GBS polypeptides of the invention.

Another diagnostic method for the detection of streptococcus in a patient comprises:

- a) labeling an antibody reactive with a polypeptide of the invention or fragment thereof with a detectable label;
- b) administering the labeled antibody or labeled fragment to the patient; and
- c) detecting specifically bound labeled antibody or labeled fragment in the patient which indicates the presence of streptococcus.

A further aspect of the invention is the use of the GBS

polypeptides of the invention as immunogens for the production of specific antibodies for the diagnosis and in particular the treatment of streptococcus infection. Suitable antibodies may be determined using appropriate screening methods, for example by measuring the ability of a particular antibody to passively protect against streptococcus infection in a test model. One example of an animal model is the mouse model described in the examples herein. The antibody may be a whole antibody or an antigen-binding fragment thereof and may in general belong to any immunoglobulin class. The antibody or fragment may be of animal origin, specifically of mammalian origin and more specifically of murine, rat or human origin. It may be a natural antibody or a fragment thereof, or if desired, a recombinant antibody or antibody fragment. The term recombinant antibody or antibody fragment means antibody or antibody fragment which were produced using molecular biology techniques. The antibody or antibody fragments may be polyclonal, or preferably monoclonal. It may be specific for a number of epitopes associated with the GBS polypeptides but is preferably specific for one.

EXAMPLE 1 Murine model of lethal Group B Streptococcus (GBS) infection

The mouse model of GBS infection is described in detail in Lancefield et al (J Exp Med 142:165-179,1975). GBS strain C388/90 (Clinical isolate obtained in 1990 from the cephalorachidian fluid of a patient suffering from meningitis, Children's Hospital of Eastern Ontario, Ottawa, Canada) and NCS246 (National Center for Streptococcus, Provincial Laboratory of Public Health for Northern Alberta, Edmonton, Canada) were respectively serotyped as type Ia/c and type II/R.

To increase their virulence, the GBS strains C388/90 (serotype Ia/c) and NCS 246 (serotype II/R) were serially passaged through mice as described previously (Lancefield et al. J Exp Med 142:165-179, 1975). Briefly, the increase of virulence was monitored using intraperitoneal inoculations of serial dilutions of a subculture in Todd-Hewitt broth obtained from either the blood or spleen of infected mice. After the last passage, infected blood samples were used to inoculate Todd-Hewitt broth. After an incubation of 2 hours at 37°C with 7% CO₂, glycerol at a final concentration of 10% (v/v) was added to the culture. The culture was then aliquoted and stored at -80° C for use in GBS challenge experiments. The number of cfu of GBS present in these frozen samples was determined. The bacterial concentration necessary to kill 100% (LD100) of the 18 weeks old mice were determined to be 3.5X10⁵ and 1.1X10⁵ respectively for GBS strain C388/90 and NCS246, which corresponded to a significant increase in virulence for both strains. Indeed, the LD100 recorded before the passages for these two strains was higher than 10⁹ cfu.

In a bacterial challenge, a freshly thawed aliquot of a virulent GBS strain was adjusted to the appropriate bacterial concentration using Todd-Hewitt broth and 1ml was injected intraperitoneally to each female CD-1 mouse. The mice used for the passive protection experiments were 6 to 8 weeks old, while the ones used for the active protection experiments were approximately 18 weeks old at the time of the challenge. All inocula were verified by colony counts. Animals were observed for any sign of infection four times daily for the first 48 h after challenge and then daily for the next 12 days. At the end of that period, blood samples were obtained from the survivors and frozen at -20°C. The spleen obtained from each mouse that survived the challenge was cultured in order to identify any remaining GBS.

EXAMPLE 2 Immunization and protection in mice with formaldehyde killed whole GBS cells

5 Formaldehyde killed GBS whole cells were prepared according to the procedures described in Lancefield et al (J Exp Med 142:165-179,1975). Briefly, an overnight culture on sheep blood agar plates (Quelab Laboratories, Montreal, Canada) of a GBS strain was washed twice in PBS buffer (phosphate buffered-saline, pH7.2), adjusted to approximately 3×10^9 10 cfu/mL and incubated overnight in PBS containing 0.3% (v/v) formaldehyde. The killed GBS suspension was washed with PBS and kept frozen at -80°C .

15 Female CD-1 mice, 6 to 8 weeks old (Charles River, St-Constant, Québec, Canada), were injected subcutaneously three times at two weeks interval with 0.1 ml of formaldehyde killed cells of GBS strain C388/90 ($\sim 6 \times 10^7$ GBS), or 0.1 ml of PBS for the control group. On the day before 20 the immunization, AlhydrogelTM (Superfos Biosector, Frederikssund, Denmark) at a final concentration of 0.14 mg or 0.21 mg of Al, was added to these preparations and incubated overnight at 4°C with agitation. Serum samples were obtained from each mouse before the beginning of the 25 immunization protocol and two weeks after the last injection. The sera were frozen at -20°C .

Eight mice in each control group injected with PBS and the group immunized with formaldehyde killed whole cells GBS 30 strain C388/90 (Ia/c) were challenged with 1.5×10^4 cfu of GBS strain C388/90 (Ia/c) one week after the third injection. All mice immunized with the formaldehyde killed GBS whole cells survived the homologous challenge while, within 5 days after the challenge, only 4 out of the 8 mice 35 injected with PBS survived from the infection. In order to increase the mortality rate in the control groups, the

bacterial suspension had to be adjusted according to the age of the mice at the time of the bacterial challenge. In subsequent challenge experiments, when mice were older than 15 weeks, the bacterial inoculum was increased to
5 concentrations between 3.0×10^5 and 2.5×10^6 cfu.

5 Table 1 Immunization of CD1 mice with formaldehyde killed whole cells of GBS and subsequent homologous challenge [strain C388/90 (Ia/c)] and heterologous challenge [strain NCS246 (II/R)].

antigenic preparations used for immunization ¹	number of living mice 14 days after the bacterial challenge (% Survival)	
	homologous challenge: strain C388/90 (Ia/c)	heterologous challenge: strain NCS246 (II/R)
1st infection		
formaldehyde killed cells of GBS strain C388/90 (Ia/c) ²	8/8 (100) ³	n.d. ⁵
control PBS	4/8 (50)	n.d.
2nd infection		
formaldehyde killed cells of GBS strain C388/90 (Ia/c)	6/6 (100) ⁴	0/6 (0) ⁶
control PBS	2/6 (33)	0/6 (0)

¹ alhydrogel™ at a final concentration of 0.14 mg or 0.21mg of Al was used;

² approximately 6X10⁷ cfu;

10 ³ intraperitoneal challenge with 1 mL Todd-Hewitt culture medium containing GBS C388/90 (Ia/c) suspension adjusted to 1.5X10⁴ cfu;

⁴ intraperitoneal challenge with 1 mL Todd-Hewitt culture medium containing GBS C388/90 (Ia/c) suspension adjusted to 2.1X10⁶ cfu;

⁵ not done;

15 ⁶ intraperitoneal challenge with 1 mL Todd-Hewitt culture medium containing GBS NCS246 (II/R) suspension adjusted to 1.2X10⁵ cfu.

In another experiment, one group of 12 mice corresponding to a control group was injected with PBS, while a second group of 12 mice was immunized with formaldehyde killed whole cells of GBS strain C388/90 (Ia/c). Six mice from each of these two groups were challenged with 2.1X10⁶ cfu of the GBS strain C388/90 (Ia/c) (Table I). As the first challenge experiment, all mice immunized with the GBS strain C388/90 (Ia/c) survived the homologous challenge. Only two out of the 6 mice injected with PBS survived the infection.

The remaining 6 mice in both groups were then used one week later to verify whether this antigenic preparation could confer cross protection against strain NCS246 (II/R) which produce a serologically distinct capsule. None of the mice infected with this second GBS strain survived the infection. The later result suggested that most of the protective immune response induced by formaldehyde killed strain C388/90 is directed against the capsular polysaccharide and that it could be restricted to strains of that particular serotype. These results clearly indicated that this particular model of infection can be efficiently used to study the protection conferred by vaccination.

15 EXAMPLE 3 Immunization of rabbit with formaldehyde killed whole GBS cells and passive protection in mice

A New Zealand rabbit (2.5 kg, Charles River, St-Constant, Québec, Canada) was immunized with formaldehyde killed cells of GBS strain C388/90 (Ia/c) to obtain hyperimmune serum. This rabbit was injected subcutaneously three times at three weeks interval with approximately 1.5×10^9 cfu of formaldehyde killed whole cells of GBS strain C388/90 (Ia/c). Freund's complete adjuvant (Gibco BRL Life Technologies, Grand Island, New York) was used as the adjuvant for the first immunization, while Freund's incomplete adjuvant (Gibco BRL) was used for the following two injections. Serum samples were obtained before the beginning of the immunization protocol and two weeks after the last injection. The sera were frozen at -20°C .

The ability of this particular rabbit hyperimmune serum to passively protect mice against a lethal infection with GBS

was also evaluated. Intraperitoneal injection of mice with either 15 or 25 μ L of hyperimmune rabbit serum 18 hours before the challenge protected 4 out of 5 mice (80%) against the infection. Comparatively, survival rates lower than 20% were recorded for mice in the control group injected with PBS or serum obtained from a rabbit immunized with meningococcal outer membrane preparation. This result clearly indicates that the immunization of another animal species with killed GBS cells can induce the production of antibodies that can passively protect mice. This reagent will also be used to characterize clones.

Table 2 Passive protection of CD-1 mice conferred by rabbit serum obtained after immunization with formaldehyde killed group B whole streptococci (strain C388/90 (Ia/c)) antigenic preparation

groups	number of living mice 14 days after the bacterial challenge with GBS strain C388/90 (Ia/c) ²	% survival
rabbit hyperimmune serum ² - 25 μ l	4/5	80
rabbit hyperimmune serum ¹ - 15 μ l	4/5	80
control rabbit serum - 25 μ l	1/5	20
control PBS	1/10	10

- ¹ Freund's complete adjuvant was used for first immunization, and Freund's incomplete adjuvant for the following two injections;
- ² intraperitoneal challenge with 1 ml Todd-Hewitt culture medium containing GBS C388/90 (Ia/c) suspension adjusted to 2×10^4 cfu.

EXAMPLE 4 Recombinant production of His.Tag-GBS fusion protein

The coding region of a GBS gene was amplified by PCR (DNA
5 Thermal Cycler GeneAmp PCR system 2400 Perkin Elmer, San
Jose, CA) from the genomic DNA of GBS strain C388/90 (Ia/c)
using the oligos that contained base extensions for the
addition of the restriction sites BglII (AGATCT) and HindIII
(AAGCTT), respectively. The PCR product was purified from
10 agarose gel using a Qiaex II gel extraction kit from Qiagen
(Chatsworth, CA), digested with the restriction enzymes
BglII and HindIII (Pharmacia Canada Inc Baie d'Urfe,
Canada), and extracted with phenol:chloroform before ethanol
precipitation. The pET-32b(+) vector (Novagen, Madison, WI)
15 containing the thioredoxin-His.Tag sequence was digested
with the restriction enzymes BglII and HindIII, extracted
with phenol:chloroform, and then ethanol precipitated. The
BglII-HindIII genomic DNA fragment was ligated to the BglII-
HindIII pET-32b(+) vector to create the coding sequence for
20 thioredoxin-His.Tag-GBS fusion protein whose gene was under
control of the T7 promoter. The ligated products were
transformed into *E. coli* strain XLI Blue MRF' ($\Delta(mcrA)183\Delta$
(*mcrCB-hsdSMR-mrr*)173 *endA1 supE44 thi-1 recA1 gyrA96 relA1*
lac [F'proAB lacI^qZAM15Tn10 (Tet^r)]^c) (Stratagene, La Jolla,
25 CA) according to the method of Simanis (Hanahan, D. DNA
Cloning, 1985, D.M. Glover (ed.), pp. 109-135). The
recombinant pET plasmid was purified using a Qiagen kit
(Qiagen, Chatsworth, CA) and the nucleotide sequence of the
DNA insert was verified by DNA sequencing (Taq Dye Deoxy
30 Terminator Cycle Sequencing kit, ABI, Foster City, CA). The
recombinant pET plasmid was transformed by electroporation
(Gene Pulser II apparatus, BIO-RAD Labs, Mississauga,
Canada) into *E. coli* strain AD494 (DE3) ($\Delta ara^+ leu7697$
 $\Delta lacX74 \Delta phoA PvuII phoR \Delta malF3 F'[lac^+(lacI^q) pro]$
35 *trxB::Kan* (DE3)) (Novagen, Madison, WI). In this strain of

E. coli, the T7 promoter controlling expression of the fusion protein, is specifically recognized by the T7 RNA polymerase (present on the λ DE3 prophage) whose gene is under the control of the lac promoter which is inducible by isopropyl- β -D-thio-galactopyranoside (IPTG).

The transformant AD494(DE3)/rpET was grown at 37°C with agitation at 250 rpm in LB broth (peptone 10g/L, Yeast extract 5g/L, NaCl 10g/L) containing 100 μ g of ampicillin (Sigma-Aldrich Canada Ltd., Oakville, Canada) per mL until the A_{600} reached a value of 0.6. In order to induce the production of the thioredoxin-His.Tag-GBS fusion protein, the cells were incubated for 2 additional hours in the presence of IPTG at a final concentration of 1mM. The bacterial cells were harvested by centrifugation.

The recombinant fusion protein produced by AD494(DE3)/rpET32 upon IPTG induction for 2h was partially obtained as insoluble inclusion bodies which were purified from endogenous *E. coli* proteins by the isolation of insoluble aggregates (Gerlach, G.F. et al 1992, Infect. Immun. 60:892). Induced cells from a 500 mL culture were resuspended in 20 mL of 25% sucrose-50mM Tris-HCl buffer (pH8.0) and frozen at -70°C. Lysis of cells in thawed suspension was achieved by the addition of 5mL of a solution of lysozyme (10mg/mL) in 250mM Tris-HCl buffer (pH8.0) followed by an incubation of 10 to 15 min on ice, and the addition of 150mL of detergent mix (5 parts of 20mM Tris-HCl buffer [pH7.4]-300mM NaCl-2% deoxycholic acid-2% Nonidet P-40 and 4 parts of 100mM Tris-HCl buffer [pH8]-50mM EDTA-2% Triton X-100) followed by 5 min incubation on ice. Upon sonication, protein aggregates were harvested by centrifugation for 30 min at 35,000 X g and a sample of the soluble cellular fraction was kept. The aggregated proteins were solubilized in 6M guanidine hydrochloride. The

presence of the fusion protein in both the soluble and insoluble fractions was shown by Western Blot analysis using the serum of a mouse injected with formaldehyde killed cells of GBS strain C388/90 (Ia/c) that survived a bacterial challenge with the corresponding GBS strain.

The purification of the fusion protein from the soluble fraction of IPTG-induced AD494(DE3)/rpET was done by affinity chromatography based on the properties of the His.Tag sequence (6 consecutive histidine residues) to bind to divalent cations (Ni^{2+}) immobilized on the His.Bind metal chelation resin (Novagen, Madison, WI). The purification method used are those described in the pET system Manual, 6th Edition (Novagen, Madison, WI). Briefly, the pelleted cells obtained from a 100mL culture induced with IPTG was resuspended in 4mL of Binding buffer (5mM imidazole-500mM NaCl-20mM Tris-HCl pH7.9), sonicated, and spun at 39,000 X g for 20 min to remove debris. The supernatant was filtered (0.45 μm pore size membrane) and deposited on a column of His.Bind resin equilibrated in Binding buffer. The column was then washed with 10 column volumes of Binding buffer followed by 6 column volumes of Wash buffer (20mM imidazole-500mM NaCl-20mM Tris-HCl pH7.9). The thioredoxin-His.Tag-GBS fusion protein was eluted with Elute buffer (1M imidazole-500mM NaCl-20mM Tris-HCl pH7.9). The removal of the salt and imidazole from the sample was done by dialysis against 3 X 1 liter PBS at 4°C.

The quantities of fusion protein obtained from either the soluble or insoluble cytoplasmic fractions of *E. coli* were estimated by Coomassie staining of a sodium dodecyl sulfate (SDS)-polyacrylamide gel with serial dilutions of these proteins and a bovine serum albumin standard (Pierce Chemical Co. Rockford, Ill.).

35

EXAMPLE 5 Recombinant production of GBS protein under
control of lambda P_L promoter

The DNA coding region of a GBS protein was inserted
5 downstream of the promoter λP_L into the translation vector
pURV22. This plasmid was derived from p629 (George et al,
1987, Bio/Technology 5:600) from which the coding region for
a portion of the herpes simplex virus type I (HSV-I)
glycoprotein (gD-1) was removed and the ampicillin
10 resistance gene replaced by a kanamycin cassette obtained
from the plasmid vector pUC4K (Pharmacia Biotech Canada
Inc., Baie D'Urfe, Canada). The vector contained a cassette
of the bacteriophage λ cI857 temperature sensitive repressor
gene from which the functional P_R promoter had been deleted.
15 The inactivation of the cI857 repressor by temperature
increase from the ranges of 30-37°C to 37-42°C resulted in
the induction of the gene under the control of λ P_L. The
translation of the gene was controlled by the ribosome
binding site cro followed downstream by a BglII restriction
20 site (AGATCT) and the ATG: ACTAAGGAGGTTAGATCTATG.

Restriction enzymes and T4 DNA ligase were used according to
suppliers (Pharmacia Biotech Canada Inc., Baie D'Urfe,
Canada; and New England Biolabs Ltd., Mississauga, Canada).
25 Agarose gel electrophoresis of DNA fragments was performed
as described by Sambrook et al. (Molecular cloning : A
laboratory Manual, 1989, Cold Spring Harbor Laboratory
Press, N.Y). Chromosomal DNA of the GBS bacteria was
prepared according to procedures described in Jayarao et al
30 (J. Clin. Microbiol., 1991, 29:2774). DNA amplification
reactions by polymerase chain reaction (PCR) were made using
DNA Thermal Cycler GeneAmp PCR system 2400 (Perkin Elmer,
San Jose, CA). Plasmids used for DNA sequencing were
purified using plasmid kits from Qiagen (Chatsworth, CA).
35 DNA fragments were purified from agarose gels using Qiaex II

gel extraction kits from Qiagen (Chatsworth, CA). Plasmid transformations were carried out by the method described by Hanahan (DNA Cloning, Glover (ed.) pp, 109-135, 1985). The sequencing of genomic DNA inserts in plasmids was done using
5 synthetic oligonucleotides which were synthesized by oligonucleotide synthesizer model 394 (the Perkin-Elmer Corp., Applied Biosystems Div. (ABI), Foster City, CA). The sequencing reactions were carried out by PCR using the Taq Dye Deoxy Terminator Cycle Sequencing kit (ABI, Foster City,
10 CA) and DNA electrophoresis was performed on automated DNA sequencer 373A (ABI, Foster City, CA). The assembly of the DNA sequence was performed using the program Sequencer 3.0 (Gene Codes Corporation, Ann Arbor, MI). Analysis of the DNA sequences and their predicted polypeptides was performed
15 with the program Gene Works version 2.45 (Intelligenetics, Inc., Mountain View CA).

The coding region of the GBS gene was amplified by PCR from GBS strain C388/90 (Ia/c) genomic DNA using oligos that
20 contained base extensions for the addition of restriction sites BglII (AGATCT) and XbaI (TCTAGA), respectively. The PCR product was purified from agarose gel using a Qiaex II gel extraction kit from Qiagen (Chatsworth, CA), digested with the restriction enzymes BglII and XbaI, and extracted with
25 phenol:chloroform before ethanol precipitation. The pURV22 vector was digested with the restriction enzymes BglII and XbaI, extracted with phenol:chloroform, and ethanol precipitated. The BglII-XbaI genomic DNA fragment was ligated to the BglII-XbaI pURV22 vector in which the GBS
30 gene was under the control of the λ PL promoter. The ligated products were transformed into *E. coli* strain XLI Blue MRF' (Δ (*mcrA*)183 Δ (*mcrCB-hsdSMR-mrr*)173 *endA1 supE44 thi-1 recA1 gyrA96 relA1 lac*[F' *proAB lacI^qZAM15 Tn10*(Tet^r)]^c) (Stratagene, La Jolla CA) according to the methods described
35 in Hanahan, supra. Transformants harboring plasmids with the

insert were identified by analysis of lysed cells submitted to electrophoresis on agarose gel (Sambrook et al, supra). The recombinant pURV22 plasmid was purified using a Qiagen kit (Qiagen, Chatsworth, CA) and the nucleotide sequence of the DNA insert was verified by DNA sequencing.

The transformant XLI Blue MRF'/rpURV22 was grown at 34°C with agitation at 250 rpm in LB broth containing 50µg of kanamycin per mL until the A₆₀₀ reached a value of 0.6. In order to induce the production of the fusion protein, the cells were incubated for 4 additional hours at 39°C. The bacterial cells were harvested by centrifugation , resuspended in sample buffer, boiled for 10 min and kept at -20°C.

15

EXAMPLE 6 Subcloning GBS protein gene in CMV plasmid pCMV-GH

The DNA coding region of a GBS protein was inserted in phase downstream of the human growth hormone (hGH) gene which was under the transcriptional control of the cytomegalovirus (CMV) promoter in the plasmid vector pCMV-GH (Tang et al, Nature, 1992, 356:152). The CMV promoter is non functional in E. coli cells but active upon administration of the plasmid in eukaryotic cells. The vector also incorporated the ampicillin resistance gene.

The coding region of the gene was amplified by PCR from genomic DNA of GBS strain C388/90 (Ia/c) using the oligos that contained base extensions for the addition of the restriction sites BglII (AGATCT) and HindIII (AAGCTT). The PCR product was purified from agarose gel using a Qiaex II gel extraction kit from Qiagen (Chatsworth, CA), digested with the restriction enzymes BglII and HindIII, and extracted with phenol:chloroform before ethanol precipitation. The pCMV-GH vector (Laboratory of Dr. Stephen

A. Johnston, Department of Biochemistry, The University of Texas, Dallas, Texas) containing the human growth hormone to create fusion proteins was digested with the restriction enzymes BamHI and HindIII, extracted with phenol:chloroform, and ethanol precipitated. The 1.3-kb BglIII-HindIII genomic DNA fragment was ligated to the BamHI -HindIII pCMV-GH vector to create the hGH-GBS fusion protein under the control of the CMV promoter. The ligated products were transformed into *E. coli* strain DH5 α [ϕ 80 *lacZ* Δ M15 *endA1* *recA1* *hsdR17* (^rK⁻K⁺) *supE44* *thi-1* λ ⁻ *gyrA96* *relA1* Δ (*lacZYA-argF*)U169] (Gibco BRL, Gaithersburg, MD) according to the methods described by Hanahan, *supra*. Transformants harboring plasmids with the insert were identified by analysis of lysed cells submitted to electrophoresis on agarose gel (Sambrook, J. et al , *supra*). The recombinant pCMV plasmid was purified using a Qiagen kit (Qiagen, Chatsworth, CA) and the nucleotide sequence of the DNA insert was verified by DNA sequencing.

20

EXAMPLE 7 Immunological activity of GBS protein to GBS challenge

Four groups of 12 female CD-1 mice (Charles River, St-Constant, Quebec, Canada) of 6 to 8 weeks were injected subcutaneously three times at three week intervals with 0.1mL of the following antigenic preparations: formaldehyde killed cells of GBS strain C388/90 ($\sim 6 \times 10^7$ cfu), 20 μ g of thioredoxin-His.Tag-GBS fusion protein obtained from the insoluble (inclusion bodies) or 20 μ g of the fusion protein, affinity purified (nickel column), from the soluble cytoplasmic fraction in *E. coli*, or 20 μ g of affinity purified (nickel column) thioredoxin-His.Tag control polypeptide. 20 μ g of QuilATM (Cedarlane Laboratories Ltd, Hornby, Canada)

was added to each antigenic preparation as the adjuvant. Serum samples were obtained from each mouse before immunization (PB) and on days 20 (TB1), 41 (TB2) and 54 (TB3) during the immunization protocols. Sera were frozen at -20°C.

An increase of the ELISA titers was recorded after each injection of the fusion protein indicating a good primary response and a boost of the specific humoral immune response after each of the second and third administration. At the end of the immunization period, the means of reciprocal ELISA titers was 456,145 for the group immunized with 20µg of fusion protein obtained from inclusion bodies compared to 290,133 for the group of mice immunized with the protein from soluble fraction in *E.coli*. The latter result suggests that the protein obtained from inclusion bodies could be more immunogenic than the soluble protein. Analysis of mice sera in ELISA using the affinity purified thioredoxin-His.Tag to coat plates showed that negligible antibody titers are made against the thioredoxin-His.Tag portion of the fusion protein. The reactivity of the sera from mice injected with the recombinant fusion protein was also tested by ELISA against formaldehyde killed whole cells of GBS strain C388/90. The antibodies induced by immunization with recombinant fusion protein also recognized their specific epitopes on GBS cells indicating that their conformation is close enough to the native streptococcal protein to induce cross-reactive antibodies.

To verify whether the immune response induced by immunization could protect against GBS infection, mice were challenged with 3.5×10^5 cfu of GBS strains C338/90 (Ia/c) and 1.2×10^5 cfu of strain NCS246 (II/R) the results of which are illustrated in tables 3 and 4 respectively. Mice immunized with control thioredoxin-His.Tag peptide were not protected against challenge with either GBS strain while those

immunized with formaldehyde killed C388/90 whole cells only provided protection against homologous challenge. The thioredoxin-His.Tag-GBS fusion protein of the invention protected mice from challenge with both GBS strains. Blood
5 and spleen culture of these mice did not reveal the presence of any GBS.

Table 3 Survival from GBS strain C388/90 (Ia/c) challenge¹

immunizing agent	no. mice surviving challenge	% survival
thioredoxin-His.Tag ²	1 / 6	17
formaldehyde killed C388/90 cells ³	6 / 6	100
thioredoxin-His.Tag-GBS fusion (inclusion body preparation) ⁴	6 / 6	100
thioredoxin-His.Tag-GBS fusion (cytoplasmic fraction) ⁴	6 / 6	100

- 5 ¹ intraperitoneal administration with 1 ml Todd-Hewitt culture medium adjusted to 3.5×10^5 cfu;
- ² 20 μ g administered; posterior legs paralyzed in surviving mouse; GBS detected in blood and spleen;
- ³ 6×10^7 cfu administered;
- ⁴ 20 μ g administered.

Table 4 Survival from GBS strain NCS246 (II/R) challenge¹

immunizing agent	no. mice surviving challenge	% survival
thioredoxin-His.Tag ²	0 / 6	0
formaldehyde killed C388/90 cells ³	2 / 6	34
thioredoxin-His.Tag-GBS fusion (inclusion body preparation) ²	5 / 5 ⁴	100
thioredoxin-His.Tag-GBS fusion (cytoplasmic fraction) ²	6 / 6	100

5 ¹ intraperitoneal administration with 1 ml Todd-Hewitt culture medium containing GBS NCS246(II/R) suspension adjusted to 1.2×10^5 cfu.

² 20 μ g administered;

³ 6×10^7 cfu administered;

10 ⁴ one mouse died during immunization.

EXAMPLE 8 Immunization with recombinant GBS protein confers protection against experimental GBS infection

15

This example illustrates the protection of mice against fatal GBS infection by immunization with the recombinant protein corresponding to the SEQ ID NO:39.

20 Groups of 10 female CD-1 mice (Charles River) were immunized subcutaneously three times at three-week intervals with 20 μ g of recombinant protein purified from E. coli strain BLR (Novagen) harboring the recombinant pURV22 plasmid vector containing the GBS gene corresponding to SEQ ID NO:42 in
 25 presence of 20 μ g of QuilATM adjuvant (Cedarlane Laboratories Ltd, Hornby, Canada) or, as control, with

QuilA™ adjuvant alone in PBS. Blood samples were collected from the orbital sinus on day 1, 22 and 43 prior to each immunization and fourteen days (day 57) following the third injection. One week later the mice were challenged with

5 approximately 10^4 to 10^6 CFU of various virulent GBS strains.

Samples of the GBS challenge inoculum were plated on TSA/5% sheep blood agar plates to determine the CFU and to verify the challenge dose. Deaths were recorded for a period of 14 days and on day 14 post-challenge, the surviving mice were
10 sacrificed and blood and spleen were tested for the presence of GBS organisms. The survival data are shown in table 5.

Prechallenge sera were analyzed for the presence of antibodies reactive with GBS by standard immunoassays. Elisa
15 and immunoblot analyses indicated that immunization with recombinant GBS protein produced in *E. coli* elicited antibodies reactive with both, recombinant and native GBS protein. Antibody responses to GBS are described in Example
9.

20

Table 5. Ability of recombinant GBS protein corresponding to SEQ ID NO: 39 to elicit protection against 8 diverse GBS challenge strains

5

Immunogen	Challenge strain		No. alive:	No. dead ¹
	Designation	Type		
rGBS protein	C388/90	Ia/c	8 : 2	(P<0.0001)
none			0 : 10	
rGBS protein	NCS 246	II/R	10 : 0	(P=0.0012)
none			3 : 7	
rGBS protein	ATCC12401	Ib	10 : 0	(P=0.001)
none			3 : 7	
rGBS protein	NCS 535	V	10 : 0	(P=0.01)
none			5 : 5	
rGBS protein	NCS 9842	VI	10 : 0	(P<0.0001)
none			0 : 10	
rGBS protein	NCS 915	III	7 : 3	(P=0.0007) ²
NCS 915-F ³			1 : 9	
none			4 : 6	
rGBS protein	NCS 954	III/R	7 : 3	(P=0.002)
NCS 954-F			4 : 6	
none			1 : 9	
rGBS protein	COH1	III	4 : 6	(P=0.0004)
COH1-F			3 : 7	
none			0 : 10	

10 ¹ Groups of 10 mice per group were used, the number of mice surviving to infection and the number of dead mice are indicated. The survival curves corresponding to recombinant GBS protein-immunized animals were compared to the survival curves corresponding to mock-immunized animals using the log-rank test for nonparametric analysis.

² Comparison analysis to NCS915-F-immunized animals.

15 ³ Animals were immunized with formaldehyde-killed GBS in presence of QuilA™ adjuvant.

20 All hemocultures from surviving mice were negative at day 14 post-challenge. Spleen cultures from surviving mice were negative except for few mice from experiment MB-11.

EXAMPLE 9 Vaccination with the recombinant GBS protein
elicits an immune response to GBS

5 Groups of 10 female CD-1 mice were immunized subcutaneously
with recombinant GBS protein corresponding to SEQ ID NO:39
as described in Example 8. In order to assess the antibody
response to native GBS protein, sera from blood samples
collected prior each immunization and fourteen days after
the third immunization were tested for antibody reactive
10 with GBS cells by ELISA using plates coated with
formaldehyde-killed GBS cells from type III strain NCS 954,
type Ib strain ATCC12401, type V strain NCS 535 or type VI
strain NCS 9842. The specificity of the raised antibodies
for GBS protein was confirmed by Western blot analyses to
15 GBS cell extracts and purified recombinant antigens. The
results shown in Figure 10 clearly demonstrate that animals
respond strongly to recombinant GBS protein used as
immunogens with median reciprocal antibody titers varying
between 12000 and 128000, for sera collected after the third
20 immunization, depending of the coating antigen. All
preimmune sera were negative when tested at a dilution of
1 :100. GBS-reactive antibodies were detectable in the sera
of each animal after a single injection of recombinant GBS
protein.

25

Example 10 Antigenic conservation of the GBS protein of the present invention

5 Monoclonal antibodies (MAbs) specific to the GBS protein of the present invention were used to demonstrate that this surface antigen is produced by all GBS and that it is also antigenically highly conserved.

10 A collection of 68 GBS isolates was used to evaluate the reactivity of the GBS-specific MAbs. These strains were obtained from the National Center for Streptococcus, Provincial Laboratory of Public Health for Northern Alberta, Canada; Centre Hospitalier Universitaire de Quebec, Pavillon CHUL, Quebec, Canada; American Type Culture Collection, USA; 15 Laboratoire de Sante Publique du Quebec, Canada; and Dept. of Infectious Disease, Children's Hospital and Medical Center, Seattle, USA. All eight Mabs were tested against the following panel of strains: 6 isolates of serotype Ia or Ia/c, 3 isolates of serotype Ib, 4 isolates of serotype II, 20 14 isolates of serotype III, 2 isolates of serotype IV, 2 isolates of serotype V, 2 isolates of serotype VI, 2 isolates of serotype VII, 1 isolate of serotype VIII, 10 isolates that were not serotyped and 3 bovine *S. agalactiae* strains. MAb 3A2 was also reacted with additional GBS: 9 25 isolates of serotype Ia/c and 10 isolates of serotype V. The strains were grown overnight on blood agar plates at 37°C in an atmosphere of 5% CO₂. Cultures were stored at - 70°C in heart infusion broth with 20% (v/v) glycerol.

30 To obtain the GBS protein-specific MAbs, mice were immunized three times at three-week intervals with 20 µg of purified recombinant GBS protein (SEQ ID NO :44) in the presence of 20% QuilA™ adjuvant. Hybridoma cell lines were generated by fusion of spleen cells recovered from immunized mice with 35 the nonsecreting SP2/O myeloma cell line as described

previously (Hamel, J. et al. 1987. J. Med. Microbiol. 23:163-170). Hybrid clone supernatants were tested for specific antibody production by ELISA using formaldehyde inactivated GBS and purified recombinant GBS protein (SEQ ID NO :39 or 44) as coating antigen, as previously described (Hamel, J. et al. 1987. J. Med. Microbiol. 23:163-170). Specific hybrid were cloned by limiting dilutions, expanded, and frozen in liquid nitrogen. Production of recombinant GBS protein was presented in Examples 4 & 5. Purified recombinant GBS protein or formaldehyde inactivated GBS were resolved by electrophoresis by using the discontinuous buffer system of Laemmli as recommended by the manufacturer and then transfer onto nitrocellulose membrane for Western immunoblotting as described previously (Martin et al. 1992. Infect. Immun. 60:2718-2725).

Western immunoblotting experiments clearly indicated that all eight MAbs recognized a protein band that corresponded to the purified recombinant GBS protein (SEQ ID NO :39). These MAbs also reacted with a protein band present in every GBS isolates tested so far. The reactivity of these GBS-specific MAbs are presented in Table 6. Each MAb reacted well with all 46 GBS. In addition, these MAbs also recognized the 3 *S. agalactiae* strains of bovine origin that were tested. MAb 3A2 also recognized nineteen GBS; 9 isolates of serotype Ia/c and 10 of serotype V. The other MAbs were not tested against these additional strains.

These results demonstrated that the GBS protein (SEQ ID NO :39) was produced by all the 65 GBS and the three 3 *S. agalactiae* strains of bovine origin that were tested so far.

More importantly, these results clearly demonstrated that the epitopes recognized by these eight GBS-specific MAbs were widely distributed and conserved among GBS. These results also indicated that these epitopes were not

restricted to serologically related isolates since representatives of all known GBS serotypes including the major disease causing groups were tested.

- 5 In conclusion, the data presented in this example clearly demonstrated that the GBS protein of the present invention is produced by all GBS and that it is antigenically highly conserved.

10

Table 6. Reactivity of eight GBS protein-specific MAb with different *S. agalactiae* strains as evaluated by Western immunoblots.

5

Mabs	Number of each serotype of <i>s. agalactiae</i> strains recognized by the MAb.											
	Ia or Ia/c (6)	Ib (3)	II (4)	III (4)	IV (2)	V (2)	VI (2)	VII (2)	VIII (1)	NT (10) 2	TOTAL (26)	Bovine (3)
3A21	6	3	4	4	2	2	2	2	1	10	46	3
5A12	6	3	4	4	2	2	2	2	1	10	46	3
6G11	6	3	4	4	2	2	2	2	1	10	46	2
8B9	6	3	4	4	2	2	2	2	1	10	46	3
8E11	6	3	4	4	2	2	2	2	1	10	46	3
12B12	6	3	4	4	2	2	2	2	1	10	46	3
18F11	6	3	4	4	2	2	2	2	1	10	46	3
20G2	6	3	4	4	2	2	2	2	1	10	46	3

1 Nine additional strains of serotype Ia/c and 10 strains of serotype V were recognized by MAb 3A2.

2 These strains were not serotyped

4

10

15

WE CLAIM:

1. An isolated polynucleotide encoding a polypeptide having at least 70% identity to a second polypeptide having a sequence selected from the group consisting of:
SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5,
SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO:10,
SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:15,
SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19,
SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:24,
SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:29,
SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:34,
SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:39,
SEQ ID NO:40, SEQ ID NO:41 and SEQ ID NO:44 or
fragments, analogs or derivatives thereof.
2. A polynucleotide according to claim 1, wherein said polynucleotide encodes a polypeptide having at least 95% identity to the second polypeptide.
3. An isolated polynucleotide encoding a polypeptide capable of generating antibodies having binding specificity for a polypeptide having a sequence selected from the group consisting of:
SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5,
SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO:10,
SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:15,
SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19,
SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:24,
SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:29,
SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:34,
SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:39,
SEQ ID NO:40, SEQ ID NO:41 and SEQ ID NO:44 or
fragments, analogs or derivatives thereof.

4. An isolated polynucleotide that is complementary to the polynucleotide of claim 1.
5. An isolated polynucleotide that is complementary to the polynucleotide of claim 3.
6. The polynucleotide of claim 1, wherein said polynucleotide is DNA.
7. The polynucleotide of claim 3, wherein said polynucleotide is DNA.
8. The polynucleotide of claim 1, wherein said polynucleotide is RNA.
9. The polynucleotide of claim 3, wherein said polynucleotide is RNA.
10. A polynucleotide which hybridizes under stringent conditions to a second polynucleotide having a sequence selected from the group consisting of :
SEQ ID NO : 1, SEQ ID NO : 7, SEQ ID NO : 13, SEQ ID NO : 22, SEQ ID NO : 27, SEQ ID NO : 32, SEQ ID NO : 37, SEQ ID NO : 42 and SEQ ID NO : 43 or fragments, analogues or derivatives thereof.
11. A polynucleotide which hybridizes under stringent conditions to a second polynucleotide having a sequence selected from the group consisting of :
SEQ ID NO : 37, SEQ ID NO : 42 and SEQ ID NO : 43.
12. A polynucleotide according to claim 11 which hybridizes under stringent conditions to a second polynucleotide having the sequence SEQ ID NO : 37.

13. A polynucleotide according to claim 11 which hybridizes under stringent conditions to a second polynucleotide having the sequence SEQ ID NO : 42.
14. A polynucleotide according to claim 11 which hybridizes under stringent conditions to a second polynucleotide having the sequence SEQ ID NO : 43.
15. A polynucleotide according to claim 10 wherein said polynucleotide has at least 95% complementarity to the second polynucleotide.
16. A polynucleotide according to claim 11 wherein said polynucleotide has at least 95% complementarity to the second polynucleotide.
17. A vector comprising the polynucleotide of claim 1, wherein said polynucleotide is operably linked to an expression control region.
18. A vector comprising the polynucleotide of claim 3, wherein said polynucleotide is operably linked to an expression control region.
19. A host cell transfected with the vector of claim 17.
20. A host cell transfected with the vector of claim 18.
21. A process for producing a polypeptide comprising culturing a host cell according to claim 19 under conditions suitable for expression of said polypeptide.
22. A process for producing a polypeptide comprising culturing a host cell according to claim 20 under condition suitable for expression of said polypeptide.

23. An isolated polypeptide having at least 70% identity to a second polypeptide having a sequence selected from the group consisting of:
SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5,
SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO:10,
SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:15,
SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19,
SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:24,
SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:29,
SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:34,
SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:39,
SEQ ID NO:40, SEQ ID NO:41 and SEQ ID NO:44 or
fragments, analogs or derivatives thereof.
24. The isolated polypeptide of claim 23 having a sequence according to SEQ ID NO : 39.
25. The isolated polypeptide of claim 23 having a sequence according to SEQ ID NO : 44.
26. An isolated polypeptide capable of generating antibodies having binding specificity for a second polypeptide having a sequence selected from the group consisting of:
SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5,
SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO:10,
SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:15,
SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19,
SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:24,
SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:29,
SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:34,
SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:39,
SEQ ID NO:40, SEQ ID NO:41 and SEQ ID NO:44 or
fragments, analogs or derivatives thereof.

27. The isolated polypeptide of claim 26 having a sequence according to SEQ ID NO : 39.
28. The isolated polypeptide of claim 26 having a sequence according to SEQ ID NO : 44.
29. An isolated polypeptide having an amino acid sequence selected from the group consisting of:
SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5,
SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO:10,
SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:15,
SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19,
SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:24,
SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:29,
SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:34,
SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:39,
SEQ ID NO:40 and SEQ ID NO:41 or fragments, analogs or derivatives thereof.
30. The isolated polypeptide of claim 29 having an amino acid sequence according to SEQ ID NO : 39.
31. An isolated polypeptide having an amino acid sequence according to SEQ ID NO : 44.
32. An isolated polypeptide according to any one of claims 29 to 31, wherein the N-terminal Met residue is deleted.
33. An isolated polypeptide according to any one of claims 29 to 30, wherein the secretory amino acid sequence is deleted.
34. A vaccine composition comprising a polypeptide according to any one of claims 23 to 31 and a pharmaceutically acceptable carrier, diluent or adjuvant.

35. A vaccine composition comprising a polypeptide according to claim 32 and a pharmaceutically acceptable carrier, diluent or adjuvant.
36. A vaccine composition comprising a polypeptide according to claim 33 and a pharmaceutically acceptable carrier, diluent or adjuvant.
37. A method for therapeutic or prophylactic treatment of streptococcal bacterial infection in an animal susceptible to streptococcal infection comprising administering to said animal a therapeutic or prophylactic amount of a composition according to claim 34.
38. A method for therapeutic or prophylactic treatment of streptococcal bacterial infection in an animal susceptible to streptococcal infection comprising administering to said animal a therapeutic or prophylactic amount of a composition according to claim 35.
39. A method for therapeutic or prophylactic treatment of streptococcal bacterial infection in an animal susceptible to streptococcal infection comprising administering to said animal a therapeutic or prophylactic amount of a composition according to claim 36.
40. A method according to any one of claims 37 to 39, wherein said animal is a bovine.
41. A method according to any one of claims 37 to 39, wherein said animal is a human.

42. A method according to any one of claims 37 to 39, wherein said bacterial infection is selected from the group consisting of group A streptococcus and group B streptococcus.
43. A method according to claim 42, wherein said bacterial infection is group B streptococcus.
44. Use of a vaccine composition according to claim 34 for the therapeutic or prophylactic treatment of streptococcal bacterial infection in an animal susceptible to or infected with streptococcal infection comprising administering to said animal a therapeutic or prophylactic amount of the composition.
45. Use of a vaccine composition according to any one of claims 35 to 36 for the therapeutic or prophylactic treatment of streptococcal bacterial infection in an animal susceptible to or infected with streptococcal infection comprising administering to said animal a therapeutic or prophylactic amount of the composition.
46. Use of a vaccine composition according to any one claims 23 to 31 for the manufacture of a vaccine for the therapeutic or prophylactic treatment of streptococcal bacterial infection in an animal susceptible to streptococcal infection comprising administering to said animal a therapeutic or prophylactic amount of the composition.
47. Use of a vaccine composition according to claim 32 for the manufacture of a vaccine for the therapeutic or

prophylactic treatment of streptococcal bacterial infection in an animal susceptible to streptococcal infection comprising administering to said animal a therapeutic or prophylactic amount of the composition.

48. Use of a vaccine composition according to claim 33 for the manufacture of a vaccine for the therapeutic or prophylactic treatment of streptococcal bacterial infection in an animal susceptible to streptococcal infection comprising administering to said animal a therapeutic or prophylactic amount of the composition.

TATCTGGCAA AGAGCCAGCT AATCGTTTTA GTTGGGCTAA AAATAAATTA TTAATCAATG 60
 S G K E P A N R F S W A K N K L L I N G
 ---->

GATTCATTGC AACTCTAGCA GCAACTATCT TATTTTTTGC AGTTCAATTC ATAGGTCTTA 120
 F I A T L A A T I L F F A V Q F I G L K

AACCAGATTA CCCTGGAAAA ACCTACTTTA TTATCCTATT GACAGCATGG ACTTTGATGG 180
 P D Y P G K T Y F I I L L T A W T L M A

CATTAGTAAC TGCTTTAGTG GGATGGGATA ATAGGTATGG TTCCTTCTTG TCGTTATTAA 240
 L V T A L V G W D N R Y G S F L S L L I

TATTATTATT CCAGCTTGGT TCAAGCGCAG GAACTTACCC AATAGAATTG AGTCCTAAGT 300
 L L F Q L G S S A G T Y P I E L S P K F

TCTTTCAAAC AATTCAACCA TTTTTACCGA TGACTTACTC TGTTTCAGGA TTAAGAGAGA 360
 F Q T I Q P F L P M T Y S V S G L R E T

CCATCTCGTT GACGGGAGAC GTTAACCATC AATGGAGAAT GCTAGTAATC TTTTLAGTAT 420
 I S L T G D V N H Q W R M L V I F L V S

CATCGATGAT ACTTGCTCTT CTTATTTATC GTAAACAAGA AGATTAATAG AAAGTATCTA 480
 S M I L A L L I Y R K Q E D

GTGATAGACT AACAGTATGA TATGGTATGT CAAAGTATTT AGGAGGAGAA GATATGTCTA 540
 M S T
 |---->

CTTTAACAAT AATTATTGCA ACATTAACCTG CTTTGGAACA TTTTATATTT ATGTATTTGG 600
 L T I I I A T L T A L E H F Y I M Y L E

AGACGTTAGC CACCCAGTCA AATATGACTG GGAAGATTTT TAGTATGTCT AAAGAAGAGT 660
 T L A T Q S N M T G K I F S M S K E E L

TGTCATATTT ACCCGTTATT AAACCTTTTTA AGAATCAAGG TGTATACAAC GGCTTGATTG 720
 S Y L P V I K L F K N Q G V Y N G L I G

GCCTATTCCT CTTTTATGGG TTATATATTT CACAGAATCA AGAAATTGTA GCTGTTTTTT 780
 L F L L Y G L Y I S Q N Q E I V A V F L

TAATCAATGT ATTGCTAGTT GCTATTTATG GTGCTTTGAC AGTTGATAAA AAAATCTTAT 840
 I N V L L V A I Y G A L T V D K K I L L

TAAAACAGGG TGGTTTACCT ATATTAGCTC TTTTAACATT CTTATTTTAA TACTACTTAG 900
 K Q G G L P I L A L L T F L F

CCGTTTCGATT TAGTTGAACG GCTTTTAGTA ATCATTTTTT TCTCATAATA CAGGTAGTTT 960

AAGTAATTTG TCTTTAAAAA TAGTATAATA TAACTACGAA TTCAAAGAGA GGTGACTTTG 1020

ATTATGACTG AGAACTGGTT ACATACTAAA GATGGTTCAG ATATTTATTA TCGTGTCTGTT 1080
 M T E N W L H T K D G S D I Y Y R V V
 |---->

GGTCAAGGTC AACCGATTGT TTTTTTACAT GGCAATAGCT TAAGTAGTCG CTATTTTGAT 1140
 G Q G Q P I V F L H G N S L S S R Y F D

AAGCAAATAG CATATTTTTC TAAGTATTAC CAAGTTATTG TTATGGATAG TAGAGGGCAT 1200
 K Q I A Y F S K Y Y Q V I V M D S R G H

GGCAAAGTC ATGCAAAGCT AAATACCATT AGTTTCAGGC AAATAGCAGT TGACTTAAAG 1260
 G K S H A K L N T I S F R Q I A V D L K

GATATCTTAG TTCATTTAGA GATTGATAAA GTTATATTGG TAGGCCATAG CGATGGTGCC 1320
D I L V H L E I D K V I L V G H S D G A

AATTTAGCTT TAGTTTTTCA AACGATGTTT CCAGGTATGG TTAGAGGGCT TTTGCTTAAT 1380
N L A L V F Q T M F P G M V R G L L L N

TCAGGGAACC TGACTATTCA TGGTCAGCGA TGGTGGGATA TTCTTTTAGT AAGGATTGCC 1440
S G N L T I H G Q R W W D I L L V R I A

TATAAATTCC TTCACTATTT AGGGAAACTC TTTCCGTATA TGAGGCAAAA AGCTCAAGTT 1500
Y K F L H Y L G K L F P Y M R Q K A Q V

ATTTGCTTA TGTTGGAGGA TTTGAAGATT AGTCCAGCTG ATTTACAGCA TGTGTCAACT 1560
I S L M L E D L K I S P A D L Q H V S T

CCTGTAATGG TTTTGGTTGG AAATAAGGAC ATAATTAAGT TAAATCATT TAAGAAACTT 1620
P V M V L V G N K D I I K L N H S K K L

GCTTCTTATT TTCCAAGGGG GGAGTTTTAT TCTTTAGTTG GCTTTGGGCA TCACATTATT 1680
A S Y F P R G E F Y S L V G F G H H I I

AAGCAAGATT CCCATGTTTT TAATATTATT GCAAAAAAGT TTATCAACGA TACGTTGAAA 1740
K Q D S H V F N I I A K K F I N D T L K

GGAGAAATTG TTGAAAAAGC TAATTGAAAA AGTCAAATCA CTGACTTCTG TGATTAAAT 1800
G E I V E K A N

TGTATTTTTT ATATCTGTTT TAGTGCTTAT TATTGTTGAA ATGATTCATT TGAAACGAAC 1860
M I H L K R T
|----->

TATTTCTGTT GAGCAACTAA AGAGTGTTTT TGGGCAATTA TCTCCAATGA ATCTTTTCTT 1920
I S V E Q L K S V F G Q L S P M N L F L

AATTATCCTT GTGGGGGTTA TCGCTGTCTT ACCGACAACC GGATATGACT TTGTACTGAA 1980
I I L V G V I A V L P T T G Y D F V L N

TGGACTTTTA CGTACAGATA AAAGCAAAAG GTATATTTTA CAGACTAGTT GGTGTATCAA 2040
G L L R T D K S K R Y I L Q T S W C I N

CACTTTAAT AACTTGTCAG GATTCGGTGG CTTAATCGAT ATTGGGTTGC GCATGGCTTT 2100
T F N N L S G F G G L I D I G L R M A F

TTATGGTAAA AAAGGTCAAG AGAAGAGTGA CCTAAGAGAA GTGACTCGTT TTTTACCCTA 2160
Y G K K G Q E K S D L R E V T R F L P Y

TCTATTTCT GGTCTGTCAT TTATTAGTGT GATTGCCTTA ATCATGAGCC ATATTTTTCA 2220
L I S G L S F I S V I A L I M S H I F H

TGCCAAAGCT AGTGTGATT ACTATTATTT GGTATTAATT GGTGCTAGTA TGTATTTTCC 2280
A K A S V D Y Y Y L V L I G A S M Y F P

TGTTATTTAT TGGATTTCTG GTCATAAAGG AAGCCATTAT TTCGGAGATA TGCCATCTAG 2340
V I Y W I S G H K G S H Y F G D M P S S

TACTCGTATA AAATTAGGTG TTGTTTCTTT TTTTGAATGG GGATGTGCGG CCGCAGCATT 2400
T R I K L G V V S F F E W G C A A A A F

TATAATTATC GGTATTATTA TGGGCATTCA TCTACCAGTT TATAAAATTT TACCACTATT 2460
I I I G Y L M G I H L P V Y K I L P L F

TTGTATTGGT TGTGCCGTCG GGATTGTATC CCTTATTCCC GGTGGATTAG GAAGTTTTGA 2520
 C I G C A V G I V S L I P G G L G S F E
 ATTAGTTCTA TTTACAGGGT TTGCTGCCGA GGGACTACCT AAAGAACTG TGGTTGCATG 2580
 L V L F T G F A A E G L P K E T V V A W
 GTTATTACTT TATCGTTTAG CCTACTATAT TATTCCATTC TTTGCAGGTA TCTATTTCTT 2640
 L L L Y R L A Y Y I I P F F A G I Y F F
 TATCCATTAT TTAGGTAGTC AAATAAATCA ACGTTATGAA AATGTCCCGA AAGAGTTAGT 2700
 I H Y L G S Q I N Q R Y E N V P K E L V
 ATCAACTGTT CTACAAACCA TGGTGAGCCA TTTGATGCGT ATTTTAGGTG CATTCTTAAT 2760
 S T V L Q T M V S H L M R I L G A F L I
 |---->
 ATTTCAACA GCATTTTTTG AAAATATTAC TTATATTATG TGGTTGCAGA AGCTAGGCTT 2820
 F S T A F F E N I T Y I M W L Q K L G L
 GGACCCATTA CAAGAACAAA TGTTATGGCA GTTTCAGGT TTATTGCTGG GGGTTTGT 2880
 D P L Q E Q M L W Q F P G L L L G V C F
 TATTCTCTTA GCTAGAATA TTTGATCAAAA AGTGAAAAAT GCTTTTCCAA TTGCTATTAT 2940
 I L L A R T I D Q K V K N A F P I A I I
 CTGGATTACT TTGACATTGT TTTATCTTAA TTTAGGTCAT ATTAGTTGGC GACTATCTTT 3000
 W I T L T L F Y L N L G H I S W R L S F
 CTGGTTTATT TTAATTATTG TAGGCTTATT AGTCATTAAG CCAACTCTCT ATAAAAACA 3060
 W F I L L L L G L L V I K P T L Y K K Q
 ATTTATTTAT AGCTGGGAAG AGCGTATTAA GGATGGAATC ATTATCGTTA GTTTAATGGG 3120
 F I Y S W E E R I K D G I I I V S L M G
 AGTTCTATTT TATATTGCAG GACTACTATT CCCTATCAGG GCTCATATTA CAGGTGGTAG 3180
 V L F Y I A G L L F P I R A H I T G G S
 TATGAACGC CTGCATTATA TCATAGCATG GGAGCCGATA GCATTGGCTA CGTTGATTCT 3240
 I E R L H Y I I A W E P I A L A T L I L
 TACTCTCGTT TATTTATGTT TGGTTAAGAT TTTACAAGGA AAATCTTGTC AGATTGGTGA 3300
 T L V Y L C L V K I L Q G K S C Q I G D
 TGTGTTCAAT GTGGATCGTT ATAAAAAACT ACTTCAAGCT TACGGTGGTT CTTCGGATAG 3360
 V F N V D R Y K K L L Q A Y G G S S D S
 CGGTTTAGCC TTTTAAATG ATAAAAGGCT CTAAGGTAC CAAAAAATG GAGAAGATTG 3420
 G L A F L N D K R L Y W Y Q K N G E D C
 CGTTGCGTTC CAATTTGTAA TTGTCAATAA TAAATGTCTT ATTATGGGGG AACCAGCCGG 3480
 V A F Q F V I V N N K C L I M G E P A G
 TGATGACACT TATATTCGTG AAGCTATTGA ATCGTTTATT GATGATGCTG ATAAGCTAGA 3540
 D D T Y I R E A I E S F I D D A D K L D
 CTATGACCTT GTTTTTTACA GTATTGGACA GAAGTTGACA CTAATTTTAC ATGAGTATGG 3600
 Y D L V F Y S I G Q K L T L L L H E Y G
 TTTTGACTTT ATGAAAGTTG GTGAGGATGC TTTAGTTAAT TTAGAAACGT TTAATCTTAA 3660
 F D F M K V G E D A L V N L E T F T L K

```

AGGGAATAAG TACAAACCTT TCAGAAATGC CCTAAATAGA GTTGAAAAGG ATGGTTTCTA 3720
G N K Y K P F R N A L N R V E K D G F Y
TTTCGAAGTT GTACAATCGC CACATAGTCA AGAGCTACTA AATAGTTTGG AAGAGATTTTC 3780
F E V V Q S P H S Q E L L N S L E E I S
TAATACTTGG TTAGAAGGAC GTCCTGAAAA AGGTTTCTCA CTAGGATATT TTAATAAAGA 3840
N T W L E G R P E K G F S L G Y F N K D
TTATTTCCAA CAAGCCCCAA TAGCTTTGGT AAAAAATGCT GAACACGAAG TTGTTGCTTT 3900
Y F Q Q A P I A L V K N A E H E V V A F
TGCTAATATT ATGCCAAACT ATGAAAAGAG TATTATCTCT ATTGATTTAA TGCCTCACGA 3960
A N I M P N Y E K S I I S I D L M R H D
TAAACAGAAA ATTCCGAATG GCGTTATGGA TTTCTCTTTT TTATCATTAT TCTCTTATTA 4020
K Q K I P N G V M D F L F L S L F S Y Y
TCAAGAGAAG GGATACCACT ATTTTGATTT GGGGATGGCA CCTTTATCAG GAGTTGGTCCG 4080
Q E K G Y H Y F D L G M A P L S G V G R
CGTTGAAACA AGTTTTGCTA AAGAGAGAAT GCGGTATCTT GTCTATCATT TCGGTAGTCA 4140
V E T S F A K E R M A Y L V Y H F G S H
TTTCTACTCA TTTAATGGTT TACACAAGTA TAAGAAGAAG TTTACACCAT TGTGGTCGGA 4200
F Y S F N G L H K Y K K K F T P L W S E
ACGTTATATT TCTTGTCTC GTTCGTCCTG GTTAATTTGT GCTATTTGTG CCCTATTAAT 4260
R Y I S C S R S S W L I C A I C A L L M
GGAAGATAGT AAAATTAAGA TTGTTAAATA AGCTTTATTT GGCAATTAAG AAGAGCATGT 4320
E D S K I K I V K
CATGCGACAT GCTCTTTTTA AATCATTTAA TACCATTGAT TGCTTGAATC TACTTTATAA 4380
TATGATGTGC TTTTAAATAT TGTTTAGCTA CTGTAGCTGC TGATTTATGC TTTACAGCTA 4440
CTTGGTAGTT CATTCTTGC ATTTCTTTTT CAGTGATATG ACCAGCAAGT TTATTGAGAG 4500
CTTTTTTTAC TTGA (SEQ ID NO:1) 4514

```

FIG. 1a
[clone1-dna/aa]

SGKEPANRFS WAKNKLLING FIATLAATIL FFAVQFIGLK PDYPGKTYFI 50
 ILLTAWTLMA LVTALVGWDN RYGSFLLSLLI LLFQLGSSAG TYPIELSPKF 100
 FQTIQPFLPM TYSVSGLRRET ISLTGDVNHQ WRMLVIFLVS SMILALLIYR 150
 KQED (SEQ ID NO:2) 154

FIG. 1b

MSTLTIIIAT LTALEHFYIM YLETLATQSN MTGKIFSMSK EELSYLPVIK 50
 LFKNQGVYNG LIGLFLLYGL YISQNQEIVA VFLINVLLVA IYGALTVDKK 100
 ILLKQGGLPI LALLTFLF (SEQ ID NO:3) 118

FIG. 1c

MTENWLHTKD GSDIYYRVVG QGQPIVFLHG NSLSSRYFDK QIAYFSKYYQ 50
 VIVMDSRGGH KSHAKLNTIS FRQIAVDLKD ILVHLEIDKV ILVGHSDGAN 100
 LALVFQTMFP GMVRGLLLNS GNLTIHGQRW WDILLVRIAY KFLHYLGKLF 150
 PYMRQKAQVI SLMLEDLKIS PADLQHVSTP VMVLVGNKDI IKLNHSHKLA 200
 SYFPRGEFYS LVGFGHHIHK QDSHVFNIIA KKFINDTLKG EIVEKAN 247
 (SEQ ID NO:4)

FIG. 1d

MIHLKRTISV	EQLKSVFGQL	SPMNLFLIIL	VGVI AVLPTT	GYDFVLNGLL	50
RTDKSKRYIL	QTSWCINTFN	NLSGFGGLID	IGLRMAFYGK	KGQEKSDLRE	100
VTRFLPYLIS	GLSFISVIAL	IMSHIFHAKA	SVDYYYLVLI	GASMYFPVIY	150
WISGHKGSY	FGDMPSSTRI	KLGVVSFFEW	GCAAAAFIII	GGLMGIHLPV	200
YKILPLFCIG	CAVGIVSLIP	GGLGSFELVL	FTGFAAEGLP	KETVVAVLLL	250
YRLAYYIIPF	FAGIYFFIHY	LGSQINQRYE	NVPKELVSTV	LQTMVSHLMR	300
ILGAFLIFST	AFFENITYIM	WLQKLGLDPL	QEQLWQFPG	LLLGVCFILL	350
ARTIDQKVKN	AFPIAIIWIT	LTLFYLNLGH	ISWRLSFWFI	LLLLGLLVIK	400
PTLYKKQFIY	SWEERIKDGI	IIVSLMGVLF	YIAGLLFPIR	AHITGGSIER	450
LHYIIAWEPI	ALATLILTLV	YLCLVKILQG	KSCQIGDVFN	VDRYKLLQA	500
YGGSSDSGLA	FLNDKRLYWY	QKNGEDCVAF	QFVIVNNKCL	IMGEPAGDDT	550
YIREAIESFI	DDADKLDYDL	VFYSIGQKLT	LLLHEYGFDF	MKVGEDALVN	600
LETFTLKGK	YKPFERNALNR	VEKDGIFYFEV	VQSPHSQELL	NSLEEISNTW	650
LEGRPEKGS	LGIFYFNKDYFQ	QAPIALVKNA	EHEVVAFANI	MPNYEKSIIS	700
IDLMRHKQK	IPNGVMDFLF	LSLFSYYQEK	GYHYFDLGMA	PLSGVGRVET	750
SFAKERMAYL	VYHFGSHFYS	FNGLHKKYK	FTPLWSERYI	SCSRSSWLIC	800
AICALLMEDS	KIKIVK	(SEQ ID NO:5)			816

FIG. 1e

MRILGAFLIF	STAFFENITY	IMWLQKLGLD	PLQEQLWQF	PGLLLGVCFI	50
LLARTIDQKV	KNAFPIAIW	ITLTLFYLN	GHISWRLSFW	FILLLLGLLV	100
IKPTLYKKQF	IYSWEERIKD	GIIIVSLMGV	LFYIAGLLFP	IRAHITGGS	150
ERLHYIIAWE	PIALATLILT	LVYLCLVKIL	QKSCQIGDV	FNVDRYKLL	200
QAYGGSSDSG	LAFLNDKRLY	WYQKNGEDCV	AFQFVIVNNK	CLIMGEPAGD	250
DTYIREAIES	FIDDADKLDY	DLVFYSIGQK	LTLLEHEYGF	DFMKVGEDAL	300
VNLETFTLKG	NKYKPFERNAL	NRVEKDGIFY	EVVQSPHSQE	LLNSLEEISN	350
TWLEGRPEKG	FSLGYFNKDY	FQQAPIALVK	NAEHEVVAFA	NIMPNYEKSI	400
ISIDLMRHK	QKIPNGVMDF	LFLSLFSYYQ	EKGHYFDLG	MAPLSGVGRV	450
ETSFAKERMA	YLVYHFGSHF	YSFNGLHKKY	KKFTPLWSER	YISCSRSSWL	500
ICAICALLME	DSKIKIVK	(SEQ ID NO:6)			518

FIG. 1f

```

AATTTTGATA TCGAAACAAC AACTTTTGAG GCAATGAAAA AGCACGCGTC ATTATTGGAG      60
N F D I   E T T   T F E   A M K K   H A S   L L E
---->
AAAATATCTG TTGAGCGTTC TTTTATTGAA TTTGATAAAC TTCTATTAGC ACCTTATTGG      120
K I S V   E R S   F I E   F D K L   L L A   P Y W

CGTAAAGGAA TGCTGGCACT AATAGATAGT CATGCTTTTA ATTATCTACC ATGCTTAAAA      180
R K G M   L A L   I D S   H A F N   Y L P   C L K

AATAGGGAAT TACAATTAAG CGCCTTTTTG TCCCAGTTAG ATAAAGATTT TTTATTTGAG      240
N R E L   Q L S   A F L   S Q L D   K D F   L F E

ACATCAGAAC AAGCTTGGGC ATCACTCATC TTGAGTATGG AAGTTGAACA CACAAAGACT      300
T S E Q   A W A   S L I   L S M E   V E H   T K T

TTTTTAAAAA AATGGAAGAC ATCAACTCAC TTTCAAAAAG ATGTTGAGCA TATAGTGGAT      360
F L K K   W K T   S T H   F Q K D   V E H   I V D

GTTTATCGTA TTCGTGAACA AATGGGATTG GCTAAAGAAC ATCTTTATCG TTATGGAAAA      420
V Y R I   R E Q   M G L   A K E H   L Y R   Y G K

ACTATAATAA AACAAGCGGA AGGTATTCGC AAAGCAAGAG GCTTGATGGT TGATTTTCGAA      480
T I I K   Q A E   G I R   K A R G   L M V   D F E

AAAATAGAAC AACTAGATAG TGAGTTAGCA ATCCATGATA GGCATGAGAT AGTTGTCAAT      540
K I E Q   L D S   E L A   I H D R   H E I   V V N

GGTGGCACCT TAATCAAGAA ATTAGGAATA AAACCTGGTC CACAGATGGG AGATATTATC      600
G G T L   I K K   L G I   K P G P   Q M G   D I I

TCTCAAATTG AATTAGCCAT TGTTTTAGGA CAACTGATTA ATGAAGAAGA GGCTATTTTA      660
S Q I E   L A I   V L G   Q L I N   E E E   A I L

CATTTTGTTA AGCAGTACTT GATGGATTAG AGAGGATTAT ATGAGCGATT TTTTAGTAGA      720
H F V K   Q Y L   M D                               M S D F   L V D
|----->
TGGATTGACT AAGTCGGTTG GTGATAAGAC GGTCTTTAGT AATGTTTCAT TTATCATCCA      780
G L T   K S V G   D K T   V F S   N V S F   I I H

TAGTTTAGAC CGTATTGGGA TTATTGGTGT CAATGGAACT GGAAAGACAA CACTATTAGA      840
S L D   R I G I   I G V   N G T   G K T T   L L D

TGTTATTTTCG GGTGAATTAG GTTTTGATGG TGATCGTTCC CCTTTTTTCAT CAGCTAATGA      900
V I S   G E L G   F D G   D R S   P F S S   A N D

TTATAAGATT GCTTATTTAA AACAAGAACC AGACTTTGAT GATTCTCAGA CAATTTTGGA      960
Y K I   A Y L K   Q E P   D F D   D S Q T   I L D

CACCGTACTT TCTTCTGACT TAAGAGAGAT GGCTTTAATT AAAGAATATG AATTATTGCT      1020
T V L   S S D L   R E M   A L I   K E Y E   L L L

TAATCACTAC GAAGAAAGTA AGCAATCAGC TCTAGAGAAA GTAATGGCAG AAATGGATTTC      1080
N H Y   E E S K   Q S R   L E K   V M A E   M D S

TTTAGATGCT TGGTCTATTG AGAGCGAAGT CAAAACAGTA TTATCCAAAT TAGGTATTAC      1140
L D A   W S I E   S E V   K T V   L S K L   G I T

TGATTTGCAG TTGTCGGTTG GTGAATTATC AGGAGGATTA CGAAGACGTG TTCAATTAGC      1200
D L Q   L S V G   E L S   G G L   R R R V   Q L A
    
```

GCAAGTATTA	TTAAATGATG	CAGATTTATT	GCTCTTAGAC	GAACCTACTA	ACCACTTAGA	1260
Q V L	L N D A	D L L	L L D	E P T N	H L D	
TATTGACACT	ATTGCATGGT	TAACGAATTT	TTTGA AAAAT	AGTAAAAAGA	CAGTGCTTTT	1320
I D T	I A W L	T N F	L K N	S K K T	V L F	
TATAACTCAT	GATCGTTATT	TTCTAGACAA	TGTTGCAACA	CGTATTTTTG	AATTAGATAA	1380
I T H	D R Y F	L D N	V A T	R I F E	L D K	
GGCACAGATT	ACAGAATATC	AAGGCAATTA	TCAGGATTAT	GTCCGACTTC	GTGCAGAACA	1440
A Q I	T E Y Q	G N Y	Q D Y	V R L R	A E Q	
AGACGAGCGT	GATGCTGCTA	GTTTACATAA	AAAGAAACAG	CTTTATAAAC	AGGAACTAGC	1500
D E R	D A A S	L H K	K K Q	L Y K Q	E L A	
TTGGATGCGT	ACTCAGCCAC	AAGCTCGTGC	AACGAAACAA	CAGGCTCGTA	TTAATCGTTT	1560
W M R	T Q P Q	A R A	T K Q	Q A R I	N R F	
TCAAAATCTA	AAAAACGATT	TACACCAAAC	AAGCGATACA	AGCGATTTGG	AAATGACATT	1620
Q N L	K N D L	H Q T	S D T	S D L E	M T F	
TGAAACAAGT	CGAATTGGGA	AAAAGGTTAT	TAATTTTGAA	AATGTCTCTT	TTTCTTACCC	1680
E T S	R I G K	K V I	N F E	N V S F	S Y P	
AGATAAATCT	ATCTTGAAAG	ACTTTAATTT	GTTAATTCAA	AATAAAGACC	GTATTGGCAT	1740
D K S	I L K D	F N L	L I Q	N K D R	I G I	
CGTTGGAGAT	AATGGTGTTG	GAAAGTCAAC	CTTACTTAAT	TTAATTGTTC	AAGATTTACA	1800
V G D	N G V G	K S T	L L N	L I V Q	D L Q	
GCCGGATTCTG	GGTAATGTCT	CTATTGGTGA	AACGATACGT	GTAGGTTACT	TTTCAACA	1860
P D S	G N V S	I G E	T I R	V G Y F	S Q Q	
ACTTCATAAT	ATGGATGGCT	CAAAACGTGT	TATTAATTAT	TTGCAAGAGG	TTGCAGATGA	1920
L H N	M D G S	K R V	I N Y	L Q E V	A D E	
GGTTAAACT	AGTGTCCGTA	CAACAAGTGT	GACAGAACTA	TTGGAACAAT	TTCTCTTTCC	1980
V K T	S V G T	T S V	T E L	L E Q F	L F P	
ACGTTTCGACA	CATGGAACAC	AAATTGCAAA	ATTATCAGGT	GGTGAGAAAA	AAAGACTTTA	2040
R S T	H G T Q	I A K	L S G	G E K K	R L Y	
CCTTTTAAAA	ATCCTGATTG	AAAAGCCTAA	TGTGTTACTA	CTTGATGAGC	CGACAAATGA	2100
L L K	I L I E	K P N	V L L	L D E P	T N D	
CTTAGATATT	GCTACATTAA	CTGTTCTTGA	AAATTTTTTA	CAAGGCTTTG	GTGGTCCTGT	2160
L D I	A T L T	V L E	N F L	Q G F G	G P V	
GATTACAGTT	AGTCACGATC	GTTACTTTTT	AGATAAAGTG	GCTAATAAAA	TTATTGCGTT	2220
I T V	S H D R	Y F L	D K V	A N K I	I A F	
TGAAGATAAC	GATATCCGTG	AATTTTTTGG	TAATTATACT	GATTATTTAG	ATGAAAAAGC	2280
E D N	D I R E	F F G	N Y T	D Y L D	E K A	
ATTTAATGAG	CAAATAATG	AAGTTATCAG	TAAAAAAGAG	AGTACCAAGA	CAAGTCGTGA	2340
F N E	Q N N E	V I S	K K E	S T K T	S R E	
AAAGCAAAGT	CGTAAAAGAA	TGTCTTACTT	TGAAAAACAA	GAATGGGCGA	CAATTGAAGA	2400
K Q S	R K R M	S Y F	E K Q	E W A T	I E D	
CGATATTATG	ATATTGGAAA	ATACTATCAC	TCGTATAGAA	AATGATATGC	AAACATGTGG	2460

D I M I L E N T I T R I E N D M Q T C G

TAGTGATTTT ACAAGGTTAT CTGATTTACA AAAGGAATTA GATGCAAAAA ATGAAGCACT 2520
 S D F T R L S D L Q K E L D A K N E A L

TCTAGAAAAG TATGACCGTT ATGAGTACCT TAGTGAGTTA GACACATGAT TATCCGTCGG 2580
 L E K Y D R Y E Y L S E L D T M I I R P
 |---->

ATTATTAATA ATGATGACCA AGCAGTTGCA CAATTAATTC GACAAAGTTT ACGCGCCTAT 2640
 I I K N D D Q A V A Q L I R Q S L R A Y

GATTTAGATA AACCTGATAC AGCATATTCA GACCCTCACT TAGATCATTG GACCTCATAC 2700
 D L D K P D T A Y S D P H L D H L T S Y

TACGAAAAAA TAGAGAAGTC AGGATTCTTT GTCATTGAGG AGAGAGATGA GATTATTGGC 2760
 Y E K I E K S G F F V I E E R D E I I G

TGTGGCGGCT TTGGTCCGCT GAAAAATCTA ATTGCAGAGA TGCAGAAGGT GTACATTGCA 2820
 C G G F G P L K N L I A E M Q K V Y I A

GAACGTTTCC GTGGTAAGGG GCTTGCTACT GATTTAGTGA AAATGATTGA AGTAGAAGCT 2880
 E R F R G K G L A T D L V K M I E V E A

CGAAAAATTG GGTATAGACA ACTTTATTTA GAGACAGCCA GTACTTTGAG TAGGGCAACT 2940
 R K I G Y R Q L Y L E T A S T L S R A T

GCGGTTTATA AGCATATGGG ATATTGTGCC TTATCGCAAC CAATAGCAAA TGATCAAGGT 3000
 A V Y K H M G Y C A L S Q P I A N D Q G

CATACAGCTA TGGATATTTG GATGATTAAA GATTTATAAG TTGAAAGTGG ATTAGTGAAC 3060
 H T A M D I W M I K D L

ATGGATTAAT TATTTTGAGA TAAGAGGAAA GAAAAGGAGA CATATATGGC ATATATTTGG 3120
 M A Y I W
 |---->

TCTTATTTGA AAAGGTACCC CAATTGGTTA TGGCTTGATT TACTAGGAGC TATGCTTTTT 3180
 S Y L K R Y P N W L W L D L L G A M L F

GTGACGGTTA TCCTAGGAAT GCCCACAGCC TTAGCGGGTA TGATTGATAA TGGCGTTACA 3240
 V T V I L G M P T A L A G M I D N G V T

AAAGGTGATC GGACTGGAGT TTATCTGTGG ACGTTCATCA TGTTTATATT TGTTGTACTA 3300
 K G D R T G V Y L W T F I M F I F V V L

GGTATTATTG GCGGTATTAC GATGGCTTAC GCATCTAGTC GCTTAACGAC AACAATGATT 3360
 G I I G R I T M A Y A S S R L T T T M I

AGAGATATGC GTAATGATAT GTATGCTAAG CTTCAAGAAT ACTCCCATCA TGAATATGAA 3420
 R D M R N D M Y A K L Q E Y S H H E Y E

CAGATAGGTG TATCTTCACT AGTGACACGT ATGACAAGCG AACTTTTTGT TTTGATGCAA 3480
 Q I G V S S L V T R M T S D T F V L M Q

TTTGCTGAAA TGTCTTTACG TTTAGGCCTA GTAACCTCCTA TGGTAATGAT TTTTAGCGTG 3540
 F A E M S L R L G L V T P M V M I F S V

GTTATGATAC TAATTACGAG TCCATCTTTG GCTTGGCTTG TAGCGGTTGC GATGCCTCTT 3600
 V M I L I T S P S L A W L V A V A M P L

TTGGTAGGAG TCGTTTTATA TGTAGCTATA AAAACAAAAC CTTTATCTGA AAGACAACAG 3660
 L V G V V L Y V A I K T K P L S E R Q Q

ACTATGCTTG	ATAAAATCAA	TCAATATGTT	CGTGAAAATT	TAACAGGGTT	ACGCGTTGTT	3720
T M L D	K I N	Q Y V	R E N L	T G L	R V V	
AGAGCCTTTG	CAAGAGAGAA	TTTTCAATCA	CAAAAATTC	AAGTCGCTAA	CCAACGTTAC	3780
R A F A	R E N	F Q S	Q K F Q	V A N	Q R Y	
ACAGATACTT	CAACTGGTCT	TTTTAAATTA	ACAGGGCTAA	CAGAACCACT	TTTCGTTCAA	3840
T D T S	T G L	F K L	T G L T	E P L	F V Q	
ATTATTATTG	CAATGATTGT	GGCTATCGTT	TGGTTTGCTT	TGGATCCCTT	ACAAAGAGGT	3900
I I I A	M I V	A I V	W F A L	D P L	Q R G	
GCTATTAAAA	TAGGGGATTT	AGTTGCTTTT	ATCGAATATA	GCTTCCATGC	TCTCTTTTCA	3960
A I K I	G D L	V A F	I E Y S	F H A	L F S	
TTTTTGCTAT	TTGCCAATCT	TTTTACTATG	TATCCTCGTA	TGGTGGTATC	AAGCCATCGT	4020
F L L F	A N L	F T M	Y P R M	V V S	S H R	
ATTAGAGAGG	TGATGGATAT	GCCAATCTCT	ATCAATCCTA	ATGCCGAAGG	TGTTACGGAT	4080
I R E V	M D M	P I S	I N P N	A E G	V T D	
ACGAAACTTA	AAGGGCATT	AGAATTTGAT	AATGTAACAT	TCGCTTATCC	AGGAGAAACA	4140
T K L K	G H L	E F D	N V T F	A Y P	G E T	
GAGAGTCCCG	TTTTGCATGA	TATTTCTTTT	AAAGCTAAGC	CTGGAGAAAC	AATGCTTTTT	4200
E S P V	L H D	I S F	K A K P	G E T	I A F	
ATTGGTTCAA	CAGGTTCAGG	AAAATCTTCT	CTTGTTAATT	TGATTCCACG	TTTTTATGAT	4260
I G S T	G S G	K S S	L V N L	I P R	F Y D	
GTGACACTTG	GAAAAATCTT	AGTAGATGGA	GTTGATGTAA	GAGATTATAA	CCTTAAATCA	4320
V T L G	K I L	V D G	V D V R	D Y N	L K S	
CTTCGCCAAA	AGATTGGATT	TATCCCCCAA	AAAGCTCTTT	TATTTACAGG	GACAATAGGA	4380
L R Q K	I G F	I P Q	K A L L	F T G	T I G	
GAGAATTTAA	AATATGGAAA	AGCTGATGCT	ACTATTGATG	ATCTTAGACA	AGCGGTTGAT	4440
E N L K	Y G K	A D A	T I D D	L R Q	A V D	
ATTTCTCAAG	CTAAAGAGTT	TATTGAGAGT	CACCAAGAAG	CCTTTGAAAC	GCATTTAGCT	4500
I S Q A	K E F	I E S	H Q E A	F E T	H L A	
GAAGGTGGGA	GCAATCTTTC	TGGGGGTCAA	AAACAACGGT	TATCTATTGC	TAGGGCTGTT	4560
E G G S	N L S	G G Q	K Q R L	S I A	R A V	
GTAAAGATC	CAGATTTATA	TATTTTTGAT	GATTCATTTT	CTGCTCTCGA	TTATAAGACA	4620
V K D P	D L Y	I F D	D S F S	A L D	Y K T	
GACGCTACTT	TAAGAGCGCG	TCTAAAAGAA	GTAACCGGTG	ATTCTACAGT	TTTGATAGTT	4680
D A T L	R A R	L K E	V T G D	S T V	L I V	
GCTCAAAGGG	TGGGTACGAT	TATGGATGCT	GATCAGATTA	TTGTCCTTGA	TGAAGCGGAA	4740
A Q R V	G T I	M D A	D Q I I	V L D	E G E	
ATTGTCGGTC	GTGGTACCCA	CGCTCAATTA	ATAGAAAATA	ATGCTATTTA	TCGTGAAATC	4800
I V G R	G T H	A Q L	I E N N	A I Y	R E I	
GCTGAGTCAC	AACTGAAGAA	CCAAAATTA	TCAGAAGGAG	AGTGATTGTA	TGAGAAAAAA	4860
A E S Q	L K N	Q N L	S E G E		M R K K	
					----->	

ATCTGTTTTT	TTGAGATTAT	GGTCTTACCT	AACTCGCTAC	AAAGCTACTC	TTTTCTTAGC	4920
S V F	L R L W	S Y L	T R Y	K A T L	F L A	
GATTTTTTTG	AAAGTTTTAT	CTAGTTTTAT	GAGTGTTCTG	GAGCCTTTTA	TTTAGGGTT	4980
I F L	K V L S	S F M	S V L	E P F I	L G L	
AGCGATAACA	GAGTTGACTG	CTAACCTTGT	TGATATGGCT	AAGGGAGTTT	CTGGGGCAGA	5040
A I T	E L T A	N L V	D M A	K G V S	G A E	
ATTGAACGTT	CCTTATATTG	CTGGTATTTT	GATTATTTAT	TTTTTCAGAG	GTGTTTTCTA	5100
L N V	P Y I A	G I L	I I Y	F F R G	V F Y	
TGAATTAGGT	TCTTATGGCT	CAAATT	(SEQ ID NO:7)			5126
E L G	S Y G S	N				

FIG. 2a

NFDIETTTFE	AMKKHASLLE	KISVERSFIE	FDKLLLAPYW	RKGMLALIDS	50
HAFNYLPCLK	NRELQLSAFL	SQLDKDFLFE	TSEQAWASLI	LSMEVEHTKT	100
FLKKWKTSTH	FQKDVEHIVD	VYRIREQMGL	AKEHLYRYGK	TIKQAEGER	150
KARGLMVDFE	KIEQLDSELA	IHDRHEIVVN	GGTLIKKLG	KPGPQMGDII	200
SQIELAIVLG	QLINEEEAIL	HFVKQYLM	(SEQ ID NO:8)		229

FIG. 2b

MSDFLVDGLT	KSVGDKTVFS	NVSFIIHSLD	RIGIIGVNGT	GKTTLLDVIS	50
GELGFDGDRS	PFSSANDYKI	AYLKQEPDFD	DSQTILDTVL	SSDLREMAI	100
KEYELLLNHY	EESKQSRLEK	VMAEMDSLDA	WSIESEVKT	LSKLGITDLQ	150
LSVGELSGGL	RRRVQLAQVL	LNDADLLLLD	EPTNHLDIDT	IAWLTNFLKN	200
SKKTVLFITH	DRYFLDNVAT	RIFELDKAQI	TEYQGNQDY	VRLRAEQDER	250
DAASLHKKKQ	LYKQELAWMR	TQPQARATKQ	QARINRFQNL	KNDLHQTSDT	300
SDLEMTFETS	RIGKKVINFE	NVSFSYPDKS	ILKDFNLLIQ	NKDRIGIVGD	350
NGVGKSTLLN	LIVQDLQPDS	GNVSIGETIR	VGYFSQQLHN	MDGSKRVINY	400
LQEVADDEVKT	SVGTTSVTEL	LEQFLFPRST	HGTQIAKLSG	GEKKRLLYLLK	450
ILIEKPNVLL	LDEPTNDLDI	ATLTVLENFL	QGFGGPVITV	SHDRYFLDKV	500
ANKIIAFEDN	DIREFFGNYT	DYLDEKAFNE	QNEVISKKE	STKTSREKQS	550
RKRMSYFEKQ	EWATIEDDIM	ILENTITRIE	NDMQTCGSDF	TRLSDLQKEL	600
DAKNEALLEK	YDRYEYLSL	DT	(SEQ ID NO:9)		622

FIG. 2c

MIIRPIIKND	DQAVAQLIRQ	SLRAYDLKDP	DTAYSDPHLD	HLTSYYEKIE	50
KSGFFVIEER	DEIIGCGGFG	PLKNLIAEMQ	KVYIAERFRG	KGLATDLVKM	100
IEVEARKIGY	RQLYLETAST	LSRATAVYKH	MGYCALSQPI	ANDQGHAMT	150
IWMIKDL	(SEQ ID NO:10)				157

FIG. 2d

MAYIWSYLKR	YPNWLWLDLL	GAMLFVTVIL	GMPTALAGMI	DNGVTKGDRT	50
GVYLWTFIMF	IFVVLGIIGR	ITMAYASSRL	TTMIRDMRN	DMYAKLQEYS	100
HHEYEQIGVS	SLVTRMTSDT	FVLMQFAEMS	LRLGLVTPMV	MIFSVVMILI	150
TSPSLAWLVA	VAMPLLVGVV	LYVAIKTKPL	SERQQTMLDK	INQYVRENLT	200
GLRVVRAFAR	ENFQSQKFQV	ANQRYTDTST	GLFKLTGLTE	PLFVQIIIAM	250
IVAIVWFALD	PLQRGAIKIG	DLVAFIEYSF	HALFSFLLFA	NLFTMYPRMV	300
VSSHRIREVM	DMPISINPNA	EGVTDTKLKG	HLEFDNVTFA	YPGETESPVL	350
HDISFKAKPG	ETIAFIGSTG	SGKSSLVNL	PRFYDVTLGK	ILVDGVDVRD	400
YNLKSLRQKI	GFIPQKALLF	TGTIGENLKY	GKADATIDDL	RQAVDISQAK	450
EFIESHQEAF	ETHLAEGGSN	LSGGQKQRLS	IARAVVKDPD	LYIFDDSESA	500
LDYKTDATLR	ARLKEVTGDS	TVLIVAQRVG	TIMDADQIIV	LDEGEIVGRG	550
THAQLIENNA	IYREIAESQL	KNQNLSEGE	(SEQ ID NO:11)		579

FIG. 2e

MRKKSVELRL	WSYLTRYKAT	LFLAIFLKV	SSFMSVLEPF	ILGLAITELT	50
ANLVDMAKGV	SGAELNVPYI	AGILIIYFFR	GVFYELGSYG	SN	92

(SEQ ID NO:12)

FIG. 2f


```

AATTTGGAAG TGCTCTATCA ACAGTTGAAG TAAAGGAGAT TATTAGTGAA GAAAACATAT      60
  F G S   A L S   T V E V   K E I   I S E   E N I W
  ---->
GGTTATATCG GCTCAGTTGC TGCCATTTTA CTAGCTACTC ATATTGGAAG TTACCAACTT      120
  L Y R   L S C   C H F T   S Y S   Y W K   L P T W

GGTAAGCATC ATATGGGTCT AGCAACAAAG GACAATCAGA TTGCCTATAT TGATGACAGC      180
          M G L   A T K   D N Q I   A Y I   D D S
          |---->
AAAGGTAAGG CAAAAGCCCC TAAAACAAAC AAAACGATGG ATCAAATCAG TGCTGAAGAA      240
  K G K A   K A P   K T N   K T M D   Q I S   A E E

GGCATCTCTG CTGAACAGAT CGTAGTCAAA ATTACTGACC AAGGCTATGT GACCTCACAC      300
  G I S A   E Q I   V V K   I T D Q   G Y V   T S H

GGTGACCATT ATCATTTTTA CAATGGGAAA GTTCCTTATG ATGCGATTAT TAGTGAAGAG      360
  G D H Y   H F Y   N G K   V P Y D   A I I   S E E

TTGTTGATGA CGGATCCTAA TTACCGTTTT AAACAATCAG ACGTTATCAA TGAAATCTTA      420
  L L M T   D P N   Y R F   K Q S D   V I N   E I L
          |---->
GACGGTTACG TTATTAAGT CAATGGCAAC TATTATGTTT ACCTCAAGCC AGGTAGTAAG      480
  D G Y V   I K V   N G N   Y Y V Y   L K P   G S K

CGCAAAAACA TTCGAACCAA ACAACAAATT GCTGAGCAAG TAGCCAAAGG AACTAAAGAA      540
  R K N I   R T K   Q Q I   A E Q V   A K G   T K E

GCTAAAGAAA AAGGTTTAGC TCAAGTGGCC CATCTCAGTA AAGAAGAAGT TGCGGCAGTC      600
  A K E K   G L A   Q V A   H L S K   E E V   A A V

AATGAAGCAA AAAGACAAGG ACGCTATACT ACAGACGATG GCTATATTTT TAGTCCGACA      660
  N E A K   R Q G   R Y T   T D D G   Y I F   S P T

GATATCATTG ATGATTTAGG AGATGCTTAT TTAGTACCTC ATGGTAATCA CTATCATTAT      720
  D I I D   D L G   D A Y   L V P H   G N H   Y H Y

ATTCTTAAAA AGGATTTGTC TCCAAGTGAG CTAGCTGCTG CACAAGCCTA CTGGAGTCAA      780
  I P K K   D L S   P S E   L A A A   Q A Y   W S Q

AAACAAGGTC GAGGTGCTAG ACCGTCTGAT TACCGCCCGA CACCAGCCCC AGGTCGTAGG      840
  K Q G R   G A R   P S D   Y R P T   P A P   G R R

AAAGCCCCAA TTCCTGATGT GACGCCTAAC CCTGGACAAG GTCATCAGCC AGATAACGGT      900
  K A P I   P D V   T P N   P G Q G   H Q P   D N G

GGCTATCATC CAGCGCCTCC TAGGCCAAAT GATGCGTCAC AAAACAAACA CCAAGAGAT      960
  G Y H P   A P P   R P N   D A S Q   N K H   Q R D

GAGTTTAAAG GAAAAACCTT TAAGGAACTT TTAGATCAAC TACACCGTCT TGATTTGAAA      1020
  E F K G   K T F   K E L   L D Q L   H R L   D L K

TACCGTCATG TGGAAGAAGA TGGGTTGATT TTTGAACCGA CTCAAGTGAT CAAATCAAAC      1080
  Y R H V   E E D   G L I   F E P T   Q V I   K S N

GCTTTTGGGT ATGTGGTGCC TCATGGAGAT CATTATCATA TTATCCAAG AAGTCAGTTA      1140
  A F G Y   V V P   H G D   H Y H I   I P R   S Q L

TCACCTCTTG AAATGGAATT AGCAGATCGA TACTTAGCTG GCCAAACTGA GGACAATGAC      1200
  S P L E   M E L   A D R   Y L A G   Q T E   D N D

TCAGGTTTCAG AGCACTCAAA ACCATCAGAT AAAGAAGTGA CACATACCTT TCTTGGTCAT      1260
  
```

S G S E H S K P S D K E V T H T F L G H
 CGCATCAAAG CTTACGGAAA AGGCTTAGAT GGTA AACCAT ATGATACGAG TGATGCTTAT 1320
 R I K A Y G K G L D G K P Y D T S D A Y
 GTTTTGTAGTA AAGAATCCAT TCATTCAGTG GATAAATCAG GAGTTACAGC TAAACACGGA 1380
 V F S K E S I H S V D K S G V T A K H G
 GATCATTTCC ACTATATAGG ATTTGGAGAA CTTGAACAAT ATGAGTTGGA TGAGGTCGCT 1440
 D H F H Y I G F G E L E Q Y E L D E V A
 AACTGGGTGA AAGCAAAAGG TCAAGCTGAT GAGCTTGCTG CTGCTTTGGA TCAGGAACAA 1500
 N W V K A K G Q A D E L A A A L D Q E Q
 GGCAAAGAAA AACCCTCTT TGACACTAAA AAAGTGAGTC GCAAAGTAAC AAAAGATGGT 1560
 G K E K P L F D T K K V S R K V T K D G
 AAAGTGGGCT ATATGATGCC AAAAGATGGT AAGGACTATT TCTATGCTCG TGATCAACTT 1620
 K V G Y M M P K D G K D Y F Y A R D Q L
 GATTTGACTC AGATTGCCTT TGCCGAACAA GAACTAATGC TTAAGATAA GAAGCATTAC 1680
 D L T Q I A F A E Q E L M L K D K K H Y
 CGTTATGACA TTGTTGACAC AGGTATTGAG CCACGACTTG CTGTAGATGT GTCAAGTCTG 1740
 R Y D I V D T G I E P R L A V D V S S L
 CCGATGCATG CTGGTAATGC TACTTACGAT ACTGGAAGTT CGTTTGTTAT CCCACATATT 1800
 P M H A G N A T Y D T G S S F V I P H I
 GATCATATCC ATGTCGTTCC GTATTCATGG TTGACGCGCG ATCAGATTGC AACAGTCAAG 1860
 D H I H V V P Y S W L T R D Q I A T V K
 TATGTGATGC AACACCCCGA AGTTCGTCCG GATGTATGGT CTAAGCCAGG GCATGAAGAG 1920
 Y V M Q H P E V R P D V W S K P G H E E
 TCAGGTTCCG TCATTCCAAA TGTTACGCCT CTTGATAAAC GTGCTGGTAT GCCAACTGG 1980
 S G S V I P N V T P L D K R A G M P N W
 CAAATTATCC ATTCTGCTGA AGAAGTTCAA AAAGCCCTAG CAGAAGGTCG TTTTGCAACA 2040
 Q I I H S A E E V Q K A L A E G R F A T
 CCAGACGGCT ATATTTTCGA TCCACGAGAT GTTTTGCCCA AAGAACTTT TGTATGGAAA 2100
 P D G Y I F D P R D V L A K E T F V W K
 GATGGCTCCT TTAGCATCCC AAGAGCAGAT GGCAGTTCAT TGAGAACCAT TAATAAATCT 2160
 D G S F S I P R A D G S S L R T I N K S
 GATCTATCCC AAGCTGAGTG GCAACAAGCT CAAGAGTTAT TGGCAAAGAA AAATACTGGT 2220
 D L S Q A E W Q Q A Q E L L A K K N T G
 GATGCTACTG ATACGGATAA ACCCAAAGAA AAGCAACAGG CAGATAAGAG CAATGAAAAC 2280
 D A T D T D K P K E K Q Q A D K S N E N
 CAACAGCCAA GTGAAGCCAG TAAAGAAGAA AAAGAATCAG ATGACTTTAT AGACAGTTTA 2340
 Q Q P S E A S K E E K E S D D F I D S L
 CCAGACTATG GTCTAGATAG AGCAACCCTA GAAGATCATA TCAATCAATT AGCACAAAAA 2400
 P D Y G L D R A T L E D H I N Q L A Q K
 GCTAATATCG ATCCTAAGTA TCTCATTTTC CAACCAGAAG GTGTCCAATT TTATAATAAA 2460
 A N I D P K Y L I F Q P E G V Q F Y N K

AATGGTGAAT TGGTAACTTA TGATATCAAG ACACTTCAAC AAATAAACCC TTAACCAAAA 2520
 N G E L V T Y D I K T L Q Q I N P
 GAAGATCTCA TTGTTAAAGC ACTGCTTTGT CAAAGCAAGT TACGGTGATT TTGAAGTCAT 2580
 TCTATGTAAC GAGTAGTGAT AAAAGTTGGA TAATAGCGGT TTTCTTTTGC AAAGAAATGG 2640
 TATCCATGTT AGAATAGTAA AAAAAGAGGA GGATTCTTGG ACTAATGTCA AATAAGTAGA 2700
 CAGAAAAC TGTTATTTTA TTGCGTTAAA ATAATTTTCT TCTTTCTGAT TAGGGGTTAG 2760
 .K I A N F Y N E E K Q N P T L
 TCCTAGATTA GCCGTATGTG GGTGTAATT GTTATAAAAA TTCTCAATGT ATTCAAAGCA 2820
 G L N A T H P N Y N N Y F N E I Y E F C
 GTCTAATTGA ACCTGTTTGA TATTTTGATA ATGTTTTCGG TTGATTTGTC TATGCTTTAA 2880
 D L Q V Q K I N Q Y H K R N I Q R H K L
 ATACTGAAA AATGCTTCAG TTACGGCATT ATCATAAGGA TATCCAGGAT TAGAAAAAGA 2940
 Y K F F A E T V A N D Y P Y G P N S F S
 ATGCATGATA TTGGCACTGC ACCCTAATAG TGAGACGCAA GAAAAACACT TTTAGGCAAT 3000
 H M A I
 <-----|
 CAGTTTTCTG TACTGTACAG GCGACTGGTC GTTTAATCTC TGTTGAATTC TAGTTTCATT 3060
 L K R Y Q V P S Q D N L R Q Q I R T E N
 ATAAAATGTA ATGTAATTTT TAACAATATT TGTTATACTA TCTTTGTTGT ATTTTCTCCT 3120
 Y F T I Y N K V I N T I S D K N Y K R R
 ATTATGGAAA TAAAAGGTTT CAGTCTTTAG GACGGTGTGA AACCATTCOA TACAGGCATT 3180
 N H F Y F T E T K L V T H F W E I C A N
 ATCTGCAGGT GTTCCTTTTC GAGACATTGA GCGGATAATG TCTTTTTCCG TGCAAGCCTG 3240
 D A P T G K R S M S R I I D K E T C A Q
 GTAGTAAGCC ATAGAAGTAT ACACTGAGCC TTGGTCACTG TGTAAGATTG CTCCTTTATT 3300
 Y Y A M
 <-----|
 TAGGCAATTT TAACTGATTA AGGGTGTCTA GTACAAAATC CGTGTCTTGA CAATCTGAGA 3360
 K P L K L Q N L T D L V F D T D Q C D S
 TAGTGTAGC TATAATTTCT CGGTTATAGA GATTCATAAT TGATGAGAGA TACAATTTAC 3420
 I T Y A I I E R N Y L N M I S S L Y L K
 AGTTACCGAA ATATAGGTAG GTAATATCTG TTACGAGCTT TTCCTTAGGC TTATCGGCAT 3480
 C N G F Y L Y T I D T V L K E K P K D A
 GGAAATCCCG ACTCAATTTA TTATCTGTTA AATAATAAGC TTTACCCAAA TTGGGAACTT 3540
 H G D R S L K N D T L Y Y A K G L N P V
 TCTTGGTACG TGTCCGACAA AGCCAGCCAT TATTTTTTCAT GATACGATAG ACTTTCTTTG 3600
 K K T R T R C L W G N N K M I R Y V K K
 TATTAACAGT CAATCCGTGG ATTTTTTTGA GCAATCGTGT AATGGTACGA TAGCCATAAA 3660
 T N V T L G H I K K L L R T I T R Y G Y
 TAAAGTGATT CTCCATACAG AGCTGTTCAA TTAATTCAAT AAGGTCATCT TTTTTTGCGG 3720
 I F H N E M
 <-----|

CTTCTCATA C TCCTTTTTCC AACGGTAATA GGTCGACCGC TTGACCTTAA AACAGTCTAG 3780
 AATGAAAAC ATCGGGTAGT TGTTTTTATA GTCTTCCACA AGCTTGATAA GACTTACTTT 3840
 ATCGATTTCC TTATCAAGCC TCGATACTTT TTTAAGAGGT CAACCTGTAA TTGTAATTGT 3900
 I S K R I L G R Y K K L L D V Q L Q L Q
 TCCACTTCAG ACAGATGTTT CAAGCCTTTA CCGTAGGTAT ATTGCTTGCC AACACCTTGA 3960
 E V E S L H E L G K G Y T Y Q K G V G Q
 TGAAAACGAT AAAGCTCCTC GTTTTCGTAC CATTTCATCC AAGTATAGAT TTGACTATTA 4020
 H F R Y L E E N E Y W K M W T Y I Q S N
 TTTTTGATGC CTAAAGTCTC CATAATAACT CTGTTAGACT TGCCTGCTTT CTTCATATCG 4080
 N K I G L T E M I V R N S K G A K K M D
 ATGCAAGCCA GCTTAGTTTC CCATGAATAT GCTTTTTTAA CCATAATAAA ACATTCCTGT 4140
 I C A L K T E W S Y A K K V M
 TTCTAGTTTA CTAATTTCA ACAGGAGTGT TTTTCTTTG TCTCATTTTA GGGATTCAGT 4200
 GCCTATTGTT GTCATCAATT ATTTTTCTAA ATTCCCAGGA CTTAAATTGT GACCCTTGGT 4260
 CGGAATGAAA GAGAAGTGTT CCTTCAATCT TTCTTTTATT AAGTGAAAAG GCAACACTTT 4320
 TCTGTACAAC ATTTATAAAG TGTTTTTCTA GGCAATTAAT CTTTTAGTCA TTGGTGTTTG 4380
 . A I L R K T M P T Q
 GTAGTTGAGA CTACCATGAA TGCGGTGGTA ATTCCACCAA TGAACATAGT CTTTAGTCTT 4440
 Y N L S G H I R H Y N W W H V Y D K T K
 AAGAGCTAGT TCTTCCAGCA ATTGAAAGGT TTCTTGATAA ACAAATTCAA TTTTGAAAGC 4500
 L A L E E L L Q F T E Q Y V F E I K F A
 ACGATACGTA CTTTCAGCTA CGGCATTGTC ATAAGGATAA CCAGCCTGAC TAAGCGAACG 4560
 R Y T S E A V A N D Y P Y G A Q S L S R
 TGTGATTCCA AAGGCTTCCA ATATTTTCATC AATTAAGTGA TTATCAAACCT CTTTGCCACG 4620
 T I G F A E L I E D I L Q N D F E K G R
 ATCTGAATGG AACATCTTGA CTTTGGTCAG GCGGTAAGGG ATGCTTTGTA TGGCTTGCTT 4680
 D S H F M K V K T L A Y P I S Q I A Q K
 AACGAGTTCA GCGGTCTTGT GCCAACCAAG AGACAGGCCG ATGATTTTAC GGTTGTATAG 4740
 V L E A T K H W G L S L G I I E R N Y L
 GTCAATGATG AGGCAAACAT AAGCCCAACG ATTGCCTACA CGAACATAGG TTAAGTCAGT 4800
 D I I L C V Y A W R N G V R V Y T L D T
 GACTAAGGCT TGTAGTGGTC TTTCTTGCTT AAATTGCCTG TCTAAGTGGT TGGGAATAGG 4860
 V L A Q L P R E Q K F Q R D L H N P I P
 GGCTTCATTC TTGCCTCTAG AATGTGGTTT GAAGGTGGCT TTCTGATAAA CAGAAACCAA 4920
 A E N K G R S H P K F T A K Q Y V S V L
 ATTGAGTCGC TTCATAATGC GTCGAATCCG ACGACGTGAA AGTGTGATAC CTTCTGTTATT 4980
 N L R K M I R R I R R R S L T I G E N N
 CAAGCATATT TTGATTTTTC TGGATCCGTA TCTAGACTCG CTATCGAGAA AAATTCCTTT 5040
 L C I K I K R S G Y R S E S D L F I R K

```

AATAGTTTCT TCAAACCTCCG TTTCAGATAC TGACTIONCAG GCTTGATAGT AATAACTTGA 5100
  I T E E F E T E S V S E V A Q Y Y Y S S
GTGTGGCATA TTCAGCCAGC GACACATCTT TGAATGCTG TATTATCCT TATTAGCAGT 5160
  H P M N L W R C M K S I S Y K D K N A T
GATTATTTCC CTTTTGTGC CATAATCACC GCTGCTTGCT TTAGGATATC TAATT 5215
  I I E R K T G Y D G S S A K P Y R I
(SEQ ID NO:13)

```

FIG. 3a

```

FGSALSTVEV KEIISEENIW LYRLSCCHFT SYSYWKLPWT 40
(SEQ ID NO:14)

```

FIG. 3b

```

MGLATKDNQI AYIDDSKGKA KAPKTNKTMD QISAEEGISA EQIVVKITDQ 50
GYVTSHGDHY HFYNGKVPYD AIISEELMT DPNYRFKQSD VINEILDGYV 100
IKVNGNYVVY LKPGSKRKN I RTKQQIAEQV AKGTKEAKEK GLAQVAHLSK 150
EEVAAVNEAK RQGRYTTDDG YIFSPTDIID DLGDAYLVPH GNHYHYIPKK 200
DLSPSELAAA QAYWSQKQGR GARPSDYRPT PAPGRRKAPI PDVTPNPGQG 250
HQPDNGGYHP APPRPNDASQ NKHQDEFKGF KTFKELLDQL HRLDLKYRHV 300
EEDGLIFEPT QVIKSNAFGY VVPHGDHYHI IPRSQLSPLE MELADRYLAG 350
QTEDNDSGSE HSKPSDKEVT HTFLGHRIKA YGKGLDGKPY DTSDAYVFSK 400
ESIHSVDSKSG VTAKHGDHFFH YIGFGELEQY ELDEVANWVK AKGQADELAA 450
ALDQEQGKEK PLFDTKKVSR KVTKDGVKGY MMPKDGKDYF YARDQLDLTQ 500
IAFAEQELML KDKKHRYDI VDTGIEPRLA VDVSSLPMHA GNATYDTGSS 550
FVIPHIDHIH VVPYSWLTRD QIATVKYVMQ HPEVRPDVWS KPGHEESGSV 600
IPNVTPLDKR AGMPNWQIIH SAEVQKALA EGRFATPDGY IFDPRDVLAK 650
ETFVWKDGSF SIPRADGSSL RTINKSDL SQ AEWQQAQELL AKKNTGDATD 700
TDKPKEKQQA DKSNNQOPS EASKEEKESD DFIDSLPDYG LDRATLEDHI 750
NQLAQKANID PKYLIFQPEG VQFYKNNGEL VTYDIKTLQQ INP 793
(SEQ ID NO:15)

```

FIG. 3c

MTDPNYRFBKQ	SDVINEILDG	YVIKVNNGNY	VYLPKPGSKRK	NIRTKQQIAE	50
QVAKGTKEAK	EKGLAQVAHL	SKEEVAAVNE	AKRQGRYTTD	DGYIFSPTDI	100
IDDLGDAYLV	PHGNHYHYIP	KKDLSPSELA	AAQAYWSQKQ	GRGARPSDYR	150
PTPAPGRRKA	PIPDVTPNPG	QGHQPDNGGY	HPAPPRPND	SONKHQRDEF	200
KGKTFKELLD	QLHRLDLKYR	HVEEDGLIFE	PTQVIKSNAF	GYVVPBGDHY	250
HIIPRSQLSP	LEMELADRYL	AGQTEDNDSG	SEHSKPSDKE	VTHTFLGHRI	300
KAYGKGLDGG	PYDTSDAYVF	SKESIHSVVK	SGVTAKHGDH	FHYIGFGELE	350
QYELDEVANW	VKAKGQADEL	AAALDQEQGK	EKPLFDTKKV	SRKVTKDGKV	400
GYMMPKDGKD	YFYARDQLDL	TQIAFAEQEL	MLKDKKHRYR	DIVDTGIEPR	450
LAVDVSSLPM	HAGNATYDTG	SSFVIPHIDH	IHVVPYSWLT	RDQIATVKYV	500
MQHPEVRPDV	WSKPGHEESG	SVIPNVTPLD	KRAGMPNWQI	IHSAEEVQKA	550
LAEGRFATPD	GYIFDPRDVL	AKETFVWKDG	SFSIPRADGS	SLRTINKSDL	600
SQAEWQQAQE	LLAKKNTGDA	TDTDKPKEKQ	QADKSNENQQ	PSEASKEEKE	650
SDDFIDSLPD	YGLDRATLED	HINQLAQKAN	IDPKYLIFQP	EGVQFYNKNG	700
ELVTYDIKTL	QQINP	(SEQ ID NO:16)			715

FIG. 3d

MHSFSNPGYP	YDNAVTEAFF	KYLKHRQINR	KHYQNIKQVQ	LDCFEYIENF	50
YNNYNPHTAN	LGLTPNQKEE	NYFNAIK	(SEQ ID NO:17)		77

FIG. 3e

MAYYQACTEK	DIIRSMRKG	TPADNACIEW	FHTVLKTETF	YFHNRRKYNK	50
DSITNIVKNY	ITFYNETRIQ	QRLNDQSPVQ	YRKLIA	(SEQ ID NO:18)	86

FIG. 3f

MENHFIYGYR	TITRLLKkih	GLTVNTKKVY	RIMKNNGWLC	RTRTKKVPNL	50
GKAYYLTDNK	LSRDFHADKP	KEKLVTDITY	LYFGNCKLYL	SSIMNLYNRE	100
IIAYTISDCQ	DTDFVLDTLN	QLKLPK	(SEQ ID NO:19)		126

FIG. 3g

MVKKAYSWET KLACIDMKKA GKSNRVIMET LGIKNNSQIY TWMKWYENEE 50
 LYRFHQGVGK QYTYGKGLEH LSEVEQLQLQ VDLLKKYRGL IRKSIK 96
 (SEQ ID NO:20)

FIG. 3h

IRYPKASSGD YGTKREIITA NKDKYSISKM CRWLNMPHSS YYYQAVESVS 50
 ETEFEETIKR IFLDSESRYG SRKIKICLNN EGITLSRRRI RRIMKRLNLV 100
 SVYQKATFKP HSRGKNEAPI PNHLDRQFKQ ERPLQALVTD LTYVRVGNRW 150
 AYVCLIIDLY NREIIGLSLG WHKTAELVKQ AIQSIPYALT KVKMFHSDRG 200
 KEFDNQLIDE ILEAFGITRS LSQAGYPYDN AVAESTYRAF KIEFVYQETF 250
 QLLEELALKT KDYVHWWNYH RIHGSLNYQT PMTKRLIA (SEQ ID NO:21)288

FIG. 3i

AATTTGAAAG CAGAATTATC TGTAGAAGAT GAGCAATATA CAGCAACAGT TTATGGTAAA 60
 N L K A E L S V E D E Q Y T A T V Y G K
 ----->
 TCTGCTCATG GTTCAACACC ACAAGAAGGT GTTAATGGGG CGACTTATTT AGCTCTTTAT 120
 S A H G S T P Q E G V N G A T Y L A L Y
 CTAAGTCAAT TTGATTTTGA AGGTCCTGCT CGTGCTTTCT TAGATGTTAC AGCCAACATT 180
 L S Q F D F E G P A R A F L D V T A N I
 ATTCACGAAG ACTTCTCAGG TGAAAACTT GGAGTAGCTT ATGAAGATGA CTGTATGGGA 240
 I H E D F S G E K L G V A Y E D D C M G
 CCATTGAGCA TGAATGCAGG TGTCTTCCAG TTTGATGAAA CTAATGATGA TAATACTATC 300
 P L S M N A G V F Q F D E T N D D N T I
 GCTCTTAATT TCCGTTACCC ACAAGGGACA GATGCTAAAA CTATCCAAAC TAAGCTTGAG 360
 A L N F R Y P Q G T D A K T I Q T .K L E
 AAACCTAACG GAGTTGAAAA AGTGACTCTT TCTGACCATG AACACACACC AACTATGTA 420
 K L N G V E K V T L S D H E H T P H Y V
 CCTATGGACG ATGAATTAGT ATCAACCTTA CTAGCTGTCT ATGAAAAGCA AACTGGTCTT 480
 P M D D E L V S T L L A V Y E K Q T G L
 AAAGGACATG AACAGGTTAT TGGTGGTGGG ACATTTGGTC GCTTACTTGA ACGGGGTGTT 540
 K G H E Q V I G G G T F G R L L E R G V
 GCATACGGTG CCATGTTCCC AGGAGATGAA AACACTATGC ATCAAGCTAA TGAGTACATG 600
 A Y G A M F P G D E N T M H Q A N E Y M
 CCTTTAGAAA ATATTTTCCG TTCGGCTGCT ATCTACGCAG AAGCTATCTA TGAATTAATC 660

P L E N I F R S A A I Y A E A I Y E L I
 AAATAAATA ATCCTTAAAC TAAATATGTG ATCAATGATA AAGGGTGGTG AAGACATGAA 720
 K .
 AGTGTCTTTG CCTCTTTTCA TAAGGTTAGA TTTGGAGACT TTATGACTGA CTTGGAAAAA 780
 M T D L E K
 |---->
 ATTATTAAG CAATAAAAAG TGATTCACAG AATCAAAATT ATACAGAAAA TGGTATTGAT 840
 I I K A I K S D S Q N Q N Y T E N G I D
 CCTTTGTTTG CTGCTCCTAA AACAGCTAGG ATCAATATTG TTGGCCAAGC ACCTGGTTTA 900
 P L F A A P K T A R I N I V G Q A P G L
 AAAACTCAAG AAGCAAGACT CTATTGGAAG GATAAATCTG GAGATCGTCT ACGCCAGTGG 960
 K T Q E A R L Y W K D K S G D R L R Q W
 CTTGGAGTTG ATGAAGAGAC ATTTTACCAT TCTGGAAAAT TTGCTGTTTT ACCTTTAGAT 1020
 L G V D E E T F Y H S G K F A V L P L D
 TTTTATTACC CAGGCAAAGG AAAATCAGGA GATTTACCCC CTAGAAAAGG TTTTGC GGAG 1080
 F Y Y P G K G K S G D L P P R K G F A E
 AAATGGCACC CTCTTATTTT AAAAGAAATG CCTAATGTTT AATTGACCTT GCTAGTTGGT 1140
 K W H P L I L K E M P N V Q L T L L V G
 CAGTATGCTC AGAAATATTA TCTTGGAAGC TCCGCACATA AAAATCTAAC AGAAACAGTT 1200
 Q Y A Q K Y Y L G S S A H K N L T E T V
 AAAGCTTACA AAGACTATCT ACCCGATTAT TTACCCCTGG TTCACCCATC ACCGCGAAAT 1260
 K A Y K D Y L P D Y L P L V H P S P R N
 CAAATTTGGC TAAAGAAGAA TCCATGGTTT GAAAAAGATC TAATCGTTGA TTTACAAAAG 1320
 Q I W L K K N P W F E K D L I V D L Q K
 ATAGTAGCAG ATATTTTAAA AGATTAAGGA TAGGAGTTGG TATGAGAGAT AATCATCTAC 1380
 I V A D I L K D . M R D N H L H
 |---->
 ACACGTATTT TTCCTATGAT TGTCAAACGG CATTGAGGA CTATATTAAT GGTTTTACAG 1440
 T Y F S Y D C Q T A F E D Y I N G F T G
 GTGAATTTAT CACGACAGAA CATTGATT TATCAAATCC TTACACCGGT CAAGACGATG 1500
 E F I T T E H F D L S N P Y T G Q D D V
 TTCCTGATTA TAGTGCTTAT TGTCAAAAA TAGATTATCT TAATCAGAAA TATGGAAATC 1560
 P D Y S A Y C Q K I D Y L N Q K Y G N R
 GATTTAAAAA AGGAATTGAA ATCGGTTATT TTAAAGATAG GGAATCAGAT ATTTTAGATT 1620
 F K K G I E I G Y F K D R E S D I L D Y
 ATTTAAAAAA TAAAGAATTT GATTTAAAAC TATTGTCAAT CCATCATAAT GGTAGGTATG 1680
 L K N K E F D L K L L S I H H N G R Y D
 ATTATCTGCA AGAAGAAGCT CTGAAAGTAC CAACAAAGGG AGCTTTTAGC AGATTACTTT 1740
 Y L Q E E A L K V P T K G A F S R L L .
 AATCGTATGG AATTTGCCAT AGGCCGTGTG GAAGCGCACG TTTTAGCTCA CTTTGATTAT 1800
 GGTTTTCGTA AGTTAACTT AGATGTAGAA GATTTAAAAC CGTTTGAAAC GCAATTGAAG 1860
 CGCATTTCAT TAAAGATGTT ATCTAAGGGG TTAGCTTTTG AACTAAATAC CAAATCCCTT 1920

TATCTATATG GGAATGAAAA ACTTTATCGC TATGCTTTAG AGATACTCAA ACAGCTTGGT 1980
 TGTAACAAT ACTCTATAGG CTCTGACGGT CATATTCCTG AACATTTTTG TTATGAATTT 2040
 GATAGACTTC AAGGTCTGCT AAAGGACTAT CAAATTGATG AAAATCATTG GATATGAGGA 2100
 AATTTTTGAT AAAAAAGCTA GGCAATATTG CTTAGCTTTT TTGTAATGCT ATTGATAGTT 2160
 TTAGTAAAA TTTCAAAAAA ATAAAGAAAT CATTTACTTG TTGCAAGCGC TTGCGTAAAT 2220
 TGTTATGATT TTATTGGTAA CAATTCATTA AAAAAAGGAGA ATGATATGAA AAGAAAAGAC 2280
 TTATTTGGTG ATAAACAAAC TCAATACACG ATTAGAAAAGT TAAGTGTGG AGTAGCTTCA 2340
 L F G D K Q T Q Y T I R K L S V G V A S
 GTTACAACAG GGGTATGTAT TTTTCTTCAT AGTCCACAGG TATTTGCTGA AGAAGTAAGT 2400
 V T T G V C I F L H S P Q V F A E E V S
 GTTTCTCCTG CAACTACAGC GATTGCAGAG TCGAATATTA ATCAGGTTGA CAACCAACAA 2460
 V S P A T T A I A E S N I N Q V D N Q Q
 TCTACTAATT TAAAAGATGA CATAAACTCA AACTCTGAGA CGGTTGTGAC ACCCTCAGAT 2520
 S T N L K D D I N S N S E T V V T P S D
 ATGCCGGATA CCAAGCAATT AGTATCAGAT GAAACTGACA CTCAAAGGG AGTGACAGAG 2580
 M P D T K Q L V S D E T D T Q K G V T E
 CCGGATAAGG CGACAAGCCT GCTTGAAGAA AATAAAGGTC CTGTTTCAGA TAAAAATACC 2640
 P D K A T S L L E E N K G P V S D K N T
 TTAGATTTAA AAGTAGCACC ATCTACATTG CAAAATACTC CCGACAAAAC TTCTCAAGCT 2700
 L D L K V A P S T L Q N T P D K T S Q A
 ATAGGTGCTC CAAGCCCTAC CTTGAAAGTA GCTAATCAAG CTCCACGGAT TGAAAATGGT 2760
 I G A P S P T L K V A N Q A P R I E N G
 TACTTTAGGC TACATCTTAA AGAATTGCCT CAAGGTCATC CTGTAGAAAG CACTGGACTT 2820
 Y F R L H L K E L P Q G H P V E S T G L
 TGGATATGGG GAGATGTTGA TCAACCGTCT AGTAATTGGC CAAATGGTGC TATCCCTATG 2880
 W I W G D V D Q P S S N W P N G A I P M
 ACTGATGCTA AGAAAGATGA TTACGGTTAT TAIGTTGATT TTAAATTATC TGAAAAACAA 2940
 T D A K K D D Y G Y Y V D F K L S E K Q
 CGAAAACAAA TATCTTTTTT AATTAATAAC AAAGCAGGGA CAAATTTAAG CGGCGATCAT 3000
 R K Q I S F L I N N K A G T N L S G D H
 CATATTCAT TATTACGACC TGAGATGAAC CAAGTTTGGT TTGATGAAAA GTACGGTATA 3060
 H I P L L R P E M N Q V W I D E K Y G I
 CATACTTATC AACCCCTCAA AGAAGGGTAT GTCCGTATTA ACTATTTGAG TTCCTCTAGT 3120
 H T Y Q P L K E G Y V R I N Y L S S S S
 AACTATGACC ACTTATCAGC ATGGCTCTTT AAAGATGTTG CAACCCCYTC AACAACTTGG 3180
 N Y D H L S A W L F K D V A T P S T T W
 CCAGATGGTA GTAATTTTGT GAATCAAGGA CTATATGGAA GGTATATTGA TGTATCACTA 3240
 P D G S N F V N Q G L Y G R Y I D V S L

AAAACTAACG CCAAAGAGAT TGGTTTTCTA ATCTTAGATG AAAGTAAGAC AGGAGATGCA 3300
 K T N A K E I G F L I L D E S K T G D A
 GTGAAAGTTC AACCCAACGA CTATGTTTTT AGAGATTTAG CTAACCATAA CCAAATTTTT 3360
 V K V Q P N D Y V F R D L A N H N Q I F
 GTAAAAGATA AGGATCCAAA GGTTTATAAT AATCCTTATT ACATTGATCA AGTGCAGCTA 3420
 V K D K D P K V Y N N P Y Y I D Q V Q L
 AAGGATGCCC AACAAATTGA TTTAACAAGT ATTCAAGCAA GTTTTACAAC TCTAGATGGG 3480
 K D A Q Q I D L T S I Q A S F T T L D G
 GTAGATAAAA CTGAAATTTT AAAAGAATTG AAAGTGACTG ATAAAAATCA AAATGCTATA 3540
 V D K T E I L K E L K V T D K N Q N A I
 CAAATTTCTG ATATCACTCT CGATACTAGT AAATCTCTTT TAATAATCAA AGGCGACTTT 3600
 Q I S D I T L D T S K S L L I I K G D F
 AATCCTAAAC AAGGTCATTT CAACATATCT TATAATGGTA ACAATGTCAT GACAAGGCAA 3660
 N P K Q G H F N I S Y N G N N V M T R Q
 TCTTGGGAAT TTAAAGACCA ACTTTATGCT TATAGTGGA AATTAGGTGC AGTTCTCAAT 3720
 S W E F K D Q L Y A Y S G N L G A V L N
 CAAGATGGTT CAAAAGTTGA AGCCAGCCTC TGGTCACCGA GTGCTGATAG TGTCACTATG 3780
 Q D G S K V E A S L W S P S A D S V T M
 ATTATTTATG ACAAAGATAA CAAAACAGG GTTGTAGCGA CTACCCCCCT TGTGAAAAAT 3840
 I I Y D K D N Q N R V V A T T P L V K N
 AATAAGGTG TTTGGCAGAC GATACTTGAT ACTAAATTAG GTATTAAAAA CTATACTGGT 3900
 N K G V W Q T I L D T K L G I K N Y T G
 TACTATTATC TTTACGAAAT AAAAAGAGGT AAGGATAAGG TTAAGATTTT AGATCCTTAT 3960
 Y Y Y L Y E I K R G K D K V K I L D P Y
 GCAAAGTCAT TAGCAGAGTG GGATAGTAAT ACTGTTAATG ATGATATTAA AACGGCTAAA 4020
 A K S L A E W D S N T V N D D I K T A K
 GCAGCTTTTG TAAATCCAAG TCAACTTGGG CCTCAAAATT TAAGTTTTGC TAAAATTGCT 4080
 A A F V N P S Q L G P Q N L S F A K I A
 AATTTTAAAG GAAGACAAGA TGCTGTTATA TACGAAGCAC ATGTAAGAGA CTTCACTTCT 4140
 N F K G R Q D A V I Y E A H V R D F T S
 GATCGATCTT TGGATGAAA ATTAAAAAAT CAATTTGGTA CCTTTGCAGC CTTTTCAGAG 4200
 D R S L D G K L K N Q F G T F A A F S E
 AACTAGATT ATTTACAGAA ATTAGGAGTT ACACACATTC AGCTTTTACC GGTATTGAGT 4260
 K L D Y L Q K L G V T H I Q L L P V L S
 TATTTTTATG TTAATGAAAT GGATAAGTCA CGCTCAACAG CTTACACTTC CTCAGACAAT 4320
 Y F Y V N E M D K S R S T A Y T S S D N
 AATTACAATT GGGGCTATGA CCCACAGAGC TATTTTGCTC TTTCTGGGAT GTATTGAGG 4380
 N Y N W G Y D P Q S Y F A L S G M Y S E
 AAACCAAAG ATCCATCAGC ACGTATCGCC GAATTTAAAC AATTAATACA TGATATTCAT 4440
 K P K D P S A R I A E L K Q L I H D I H

AAACGTGGCA TGGGGGTTAT ACTTGATGTC GTCTATAATC ACACTGCAAA AACTTATCTC 4500
 K R G M G V I L D V V Y N H T A K T Y L
 TTTGAGGATA TAGAACCTAA TTATTATCAC TTTATGAATG AAGATGGTTC ACCAAGAGAA 4560
 F E D I E P N Y Y H F M N E D G S P R E
 AGTTTTGGAG GGGGACGTTT AGGAACCACT CATGCAATGA GTCGTCGTGT TTTGGTTGAT 4620
 S F G G G R L G T T H A M S R R V L V D
 TCCATTAAAT ATCTTACAAG TGAATTTAAA GTTGATGGTT TCCGTTTGA TATGATGGGA 4680
 S I K Y L T S E F K V D G F R F D M M G
 GATCATGATG CGGCTGCGAT TGAATTAGCT TATAAAGAAG CTAAAGCTAT TAATCCTAAT 4740
 D H D A A A I E L A Y K E A K A I N P N
 ATGATTATGA TTGGTGAGGG CTGGAGAACA TTCCAAGGCG ATCAAGGTCA GCCGGTTAAA 4800
 M I M I G E G W R T F Q G D Q G Q P V K
 CCAGCTGACC AAGATTGGAT GAAGTCAACC GATACAGTTG GCGTCTTTTC AGATGATATT 4860
 P A D Q D W M K S T D T V G V F S D D I
 CGTAATAGCT TGAAATCTGG TTTTCCAAAT GAAGGTACTC CAGCTTTCAT CACAGGTGGC 4920
 R N S L K S G F P N E G T P A F I T G G
 CCACAATCTT TACAAGGTAT TTTTAAAAAT ATCAAAGCAC AACCTGGGAA TTTTGAAGCA 4980
 P Q S L Q G I F K N I K A Q P G N F E A
 GATTCGCCAG GAGATGTGGT GCAGTATATT GCTGCACATG ATAACCTTAC CTTGCATGAT 5040
 D S P G D V V Q Y I A A H D N L T L H D
 GTGATTGCAA AATCAATT (SEQ ID NO:22) 5058
 V I A K S I .

FIG. 4a

NLKAELSVED EQYTATVYGK SAHGSTPQEG VNGATYLALY LSQDFDFEGPA 50
 RAFLDVTANI IHEDFSGEKL GVAYEDDCMG PLSMNAGVFQ FDETNDNTI 100
 ALNFRYPQGT DAKTIQTKLE KLNGVEKVTL SDHEHTPHYV PMDDELVSTL 150
 LAVYEKQTGL KGHEQVIGGG TFGRLLEGRV AYGAMFPGDE NTMHQANEYM 200
 PLENIFRSAA IYAEAIYELI K (SEQ ID NO:23) 221

FIG. 4b

MTDLEKIIKA IKSDSQNQNY TENGIDPLFA APKTARINIV GQAPGLKTQE 50
 ARLYWKDKSG DRLRQWLGV D EETFYHSGKF AVLPLDFYYP GKGKSGDLPP 100
 RKGFAEKWHP LILKEMPVQ L TLLVGQY AQ KYLLGSSAHK NLTETVKAYK 150
 DYLPDYLPV HPSRNQIWL KKNPWFEKDL IVDLQKIVAD ILKD 194
 (SEQ ID NO:24)

FIG. 4c

MRDNHLHTYF SYDCQTAFED YINGFTGEFI TTEHFDLSNP YTGQDDVPDY 50
 SAYCQKIDYL NQKYGNRFKK GIEIGYFKDR ESDILDYLKN KEFDLKLISI 100
 HHNGRYDYLQ EEALKVPTKG AFSRLL (SEQ ID NO:25) 126

FIG. 4d

MKRKDLFGDK QTQYTIRKLS VGVASVTTGV CIFLHSPQVF AEEVSVSPAT 50
 TAIAESNINQ VDNQQSTNLK DDINSNSETV VTPSDMPDTK QLVSDETDTQ 100
 KGVTEPDKAT SLLEENKGPV SDKNTLDLKV APSTLQNTPD KTSQAIGAPS 150
 PTLKVANQAP RIENGYFRLH LKELPQGHV ESTGLWIWGD VDQPSSNWPW 200
 GAIPMTDAKK DDYGYVDFK LSEKQRKQIS FLINNKAGTN LSGDHHIPLL 250
 RPEMNQVWID EKYGIHTYQP LKEGYVRINY LSSSSNYDHL SAWLFKDVAT 300
 PSTTWPDGSN FVNQGLYGRY IDVSLKTNK EIGFLILDES KTGDAVKVQP 350
 NDYVFRDLAN HNQIFVKDKD PKVYNNPYYI DQVQLKDAQQ IDLTSIQASF 400
 TTLDGVDKTE ILKELKVTDK NQNAIQISDI TLDTSKSLLI IKGDFNPKQG 450
 HFNISYNGNN VMTRQSWEFK DQLYAYSGNL GAVLNQDGSK VEASLWSPSA 500
 DSVTMIIYDK DNQNRVVATT PLVKNNKGVW QTILDTKLGI KNYTGYYYLY 550
 EIKRGKDKVK ILDPYAKSLA EWDSNTVNDD IKTAKAAFVN PSQLGPQNLS 600
 FAKIANFKGR QDAVIYEAHV RDFTSDRSLD GKLKNQFGTF AAFSEKLDYL 650
 QKLGVTHIQL LPVLSYFYVN EMDKSRSTAY TSSDNNYNWG YDPQSYFALS 700
 GMYSEKPKDP SARIAELKQL IHDIHCRGMG VILDVVYNHT AKTYLFEDIE 750
 PNYHFMNED GSPRESFGGG RLGTTTHAMSR RVLVDSIKYL TSEFKVDGFR 800
 FDMMGDHDA AIELAYKEAK AINPNMIMIG EGWRTFQGDQ GQPVKPADQD 850
 WMKSTDTVGV FSDDIRNSLK SGFPNEGTPA FITGGPQSLQ GIFKNIKAQP 900
 GNFEADSPGD VVQYIAAHDN LTLHDVIAKS I (SEQ ID NO:26) 931

FIG. 4e

```

AATTCAAAGT TTGACAGAAG GTCAACTTCG TTCTGATATC CCTGAGTTCC GTGCTGGTGA      60
  I Q S L T E G Q L R S D I P E F R A G D
  ---->
TACTGTACGT GTTCACGCTA AAGTTGTTGA AGGTA CTGCGC GAACGTATTC AGATCTTTGA      120
  T V R V H A K V V E G T R E R I Q I F E
AGGTGTTGTT ATCTCACGTA AAGGTCAAGG AATCTCAGAA ATGTACACAG TACGTAAAAT      180
  G V V I S R K G Q G I S E M Y T V R K I
TTCTGGTGGT ATCGGTGTAG AGCGTACATT CCCAATTCAC ACTCCTCGTG TTGATAAAAT      240
  S G G I G V E R T F P I H T P R V D K I
CGAAGTTGTT CGTTATGGTA AAGTACGTCG TGCTAAACTT TACTACTTAC GCGCATTGCA      300
  E V V R Y G K V R R A K L Y Y L R A L Q
AGGTAAAGCT GCACGTATTA AAGAAATCCG TCGTTAATTT TGATGATCAG ATTTTAAAAA      360
TGCTTGTTG TTTGAGGATA GTA ACTATGT TTTAAACTG GACAACCAAG ACGTAAAAAA      420
TCTGCCTGTG GGCAGTTTTT TTACTAGGTC CCCTTAGTTC AATGGATATA ACAACTCCCT      480
                . H I Y C S G
CCTAAGGAGT AATTGCTGGT TCGATTCCGG CAGGGGACAT ATTCATTGCA TGTAATAAGC      540
  G L S Y N S T R N R C P V Y E N C T F L
GGTTTAGAGC TATTTTGCCC CAAATTTCTC TGATTAAGTT TATCGTTCCT ATCTTTTTGT      600
  P K S S N Q G L N R Q N L K D N R D K Q
TCTTGTAAAT GATGTGCGTA AACTTCTAAA GTGATATTTA AATTCTCGTG ATCTAAA ACT      660
  E Q L Q H A Y V E L T I N L N E H D L V
TGAGAGATGG AAATTAGATA GCTTGCAAAT GTATGCCTGA GAGAGTGCAC TCGTACCTCG      720
  Q S I S I L Y S A F T H R L S H V R V E
CGACCAGTTA TTTTTCGGAT AGTTTTATTG ACTGCATTAT TTGAAAGTTT GTCGAATAAT      780
  R G T I K R I T K N V A N N S L K D F L
CTGTCGTTTT TATTTTTTGT AAATTCATGC AAAAAAATA ATGTATCATT GTCAATTGGT      840
  R D N K N K T F E H L F F L T D N D I P
ATATTTCTGA TACTACTTTT GTTTTTTGT GGCAGGTATC TTTGGTTGAA ATGATAATCC      900
  I N R I S S K N K T P L Y R Q N F H Y D
CAAGTTTTAT TAATTGATAA ATATTTGTTA GTGTAATCAA TATCATTAAC TGTTAAACCT      960
  W T K N I S L Y K N T Y D I D N V T L G
AAACATTCAG CGAAGCGCAT GCCAGTTTTA GCGATGAGGT ATAACGCTGC ATACGATTGA     1020
  L C E A F R M
                <----|
TGTTGTGATT TTTCTTTACA AATTTTTATC AAGCGTAAGT ATTCATTGGT TTCAAGAAAT     1080
TTTATCTCTA TTTACGCCCC TTATTTTTTG CTTTAACCTT AGTGAATAAA CAAAAATTTT     1140
TTTCTATATA TCCCTCGTGA ACAGCCATGG ATACGCAGGC TTTTACATGT ATGTTAAAAC     1200
GCTTTACTGT ATCTTGCACA TCGTTTTGAC TATAATGATT TATGACTTGT TGATATTTAG     1260

```

TGGAAGTAAT ATTGCAAAGT AATATATTTT CTATTATATG TTTATACGAT ATTCGATATT 1320
 CCCACCCGTT GTCGCGTTTA CGGAAATACG CCATTGATAT ACTCCACATT AGCTAAAGAA 1380
 CAGGGTGTTT AAGGCTACCT TGATGGAAAA GGCTCTCTTA GAGATATTTG TAAATGGTAT 1440
 GATATCTCAA GTCGCTCTGT TCTCCAAAAG TGGATAAAAC GGTATACTAG TGGTGAAGAC 1500
 TTGAAAGCCA CTAGTAGAGG ATATAGCCGT ATGAAACAAG GAAGGCAAGC CACATTTGAA 1560
 GAACGTGTAG AGATTGTTAA CTACACCATT GCCCATGGGA AAGACTATCA AGCAGCTATT 1620
 GAGAAGTTTG GTGTTTCCTA CCAACAAATT TATTCTTGGG TGCGTAAGCT TGAGAAGAAT 1680
 GGCTCACAAG GTTTGGTTGA TAGACGTGTG AAAGGGTTGG AGAGTAGGCC TGATTTAACC 1740
 GAGATTGAGC AACTTTAACT CAAGATTAAC CAATTGGAGG AACGTAATCG TCTCTTAGAA 1800
 ATCGAGGTTA GTTTACTAAA AAAGTTAGAA GACATCAAAC GAGGAAACAG ACGGTAAGAC 1860
 TAGGTAAGCA TTTAGCGGAG TTCCAAGTAA TCAAGAATTA TTACGATGAG GAATCTAATG 1920
 TGCCTATTCA GGCCTTATGC CAACTCTTGA AGGGGTCTCG TTCAGGCTAT TACAAGTGGC 1980
 TCAATCGTCA AAAACAGAT TTTGAGACAA AAAATACAAA GCTAATGGCT AAAATCAAGG 2040
 AACTTCGTAG ACTCTACAAT GGTATCTTAG GTTATCGCCG TATGACAACA TTTATTAATC 2100
 GTCAACTTGG GACAACCTAA AACAAGAAAC GGATTCGTTG ATTGATGAAC ATTCTGGGGA 2160
 TTAGTTCAGT CATTCGTCGT GTTAGCCATG CTTGTACAAA AGCTGGTGAC AGATTTTACG 2220
 AAGAAAATAT TCTTAATCGT GAATTTACAG CCACAGCTCA TAACCAGAAA TGGTGCACAG 2280
 ATGTCACCTA TCTTCAATAC GGTCTGGGAG CTAAAGCTTA TCTCAGTGCG ATTAAGACC 2340
 TGTATAACGG TTCTATTATC GCTTATGAGA TTAGTCACAA CAATGAAATC CACTTGTTAT 2400
 GAAGACCATT AAAAAGGGGC TAGAGCTCAA TCCAGGAGCC ACACCTATCA TCCATAGCGA 2460
 TTGAGGTAGT CAATATACTT CCAAAGAATA CCGTTATATC ATACAACAAG CTGGTCTGAC 2520
 CTTATCCATG TCCCGGATTG GCAAATGTAT TGATAATGCA CCAACTGAAA GTTTCTTTGG 2580
 GTTTTCAAG ACTGAGTCTT ACCACCTTAA GAAATACAAC TCTTATGATG AGTTGGTCAA 2640
 TGATGTGGCA CGTTATATCG AATTCTACAA CACACAACGT TATCAATCAA AATTAACAA 2700
 CCTGACTCCT CTAGAATTCA GGAATCAGGT TGCATAACTT ATCTTTTATT ATTTGACTGT 2760
 CTA CTGACA GGGAGCCGTT CAGATTGCTT AACCTTTCTA AATTTGCTAA AATAGCTACA 2820
 AGAAAACGAG CCATTTAATG CTTATTTCTT AACTGTCTT GCCTCAGCT CTCCTGACC 2880
 AAAAATTGAG CGTGAGGCTT TTTGTTTCAT TAAACGATGA TATTTCCATA TTCATCAGTT 2940
 TGTTTTCCGA GAGCCATCAA AGCTTCGATA AGGTCGATAA TTCCAGGAAT AAAGGTAATA 3000
 CTAAAAATA TATATAAAAA AACCTGGCCT ATTTTCTCTG CGTAAAATTT ATGCGCTCCA 3060
 ATGCCGCCA AAAGAACGTT AATAAAACAT AAACACTAT GTTAGCATAA GACTTTATTT 3120

TTACAAC TGA ATTTTCATATA AATGGATTAG AGTAAGGGAT AAAAGAAATT AGCATAGCTC 3180
 TTTTGAAAAT AAAAAAATTA ATATAATATG GAAAAAATTT TATTTTCATAA ACGTTTCATA 3240
 AAAGGTATGT AATCTAGTAT TTAGGCAACA CTATTTTGTC ACTGGTGTCT AGTAACTTAT 3300
 AGATTGATAA TTTTACTAGT AAACGTAATT CTTGCTTTA AGAGTTAAAT GTCTATTTAT 3360
 TGTAAGCTAA ATTGGGAGGT GAACTTATGT AAAATTAGAT AGGTACTGTC AAGTACGGGA 3420
 TGATTATTGA AACAGCCAGT ATGCATCATA AAATCTGTAT TGCTTAATAA CTATTTTCCTT 3480
 AACCAGACAT CAGTTCATTG TTTATCATCG CTACCCTAAG TCTAGTTTTT TCAATAGAGC 3540
 ATTAGGTAGT TTTTGATAAT AAAACTATAT AAACATGAGA ATTAGATTTT GTATTGCATT 3600
 CTTCATAATG AGTTATTGA GATTTTCCTT TGAATAAATA GATACGAAAT TCAGTAACTT 3660
 CATATATAAA CGGCTCTATC ATTGAGATAG TTTGTCAAAT GAAGAAATTT TTAATGGAAA 3720
 TAGTTTTAAA AACATTAGTT GTAGGCGATG TAAAAATATT AATCCAGTGG ATGCAATAGT 3780
 TGCGGAGTAA AAATAGAGAG GAGTAATTAG GAAGTGATAA AAAATGCTAT AGCATATATT 3840
 ACCAGAAAAA AAAATAGAAC ACTTATTATA TTTGCTATTT TAACAATTGT TCTTTCTTGC 3900
 TTGTATTCAT GTTAAACAAT AATGAAATCA AGTAATGAAA TAGAAAAGGC TTTATATGAA 3960
 M K S S N E I E K A L Y E
 |---->
 AGTTCTAATT CTTCAATATC AATTACAAA AAAGATGGTA AATATTTTAA TATTAATCAA 4020
 S S N S S I S I T K K D G K Y F N I N Q
 TTTAAGAATA TTGAAAAAAT AAAAGAGGTT GAAGAAAAAA TATTTCAATA TGATGGATTA 4080
 F K N I E K I K E V E E K I F Q Y D G L
 GCAAAATTGA AAGATCTTAA AGTAGTTAGT GGTGAGCAAA GTATAAATAG AGAAGATTTA 4140
 A K L K D L K V V S G E Q S I N R E D L
 TCTGACGAAT TTAAAAATGT TGTTTCACTA GAAGCTACAA GTAATACTAA AAGAAATCTT 4200
 S D E F K N V V S L E A T S N T K R N L
 TTATTTAGTA GTGGAGTATT TAGTTTTAAA GAAGGAAAAA ATATAGAAGA AAATGATAAG 4260
 L F S S G V F S F K E G K N I E E N D K
 AATTCAATTC TTGTTTCATGA AGAATTTGCT AAACAAAACA AACTAAAATT GGGTGATGAA 4320
 N S I L V H E E F A K Q N K L K L G D E
 ATTGATCTTG AATTACTAGA TACGGAAAAA AGTGGAAAAA TAAAAAGTCA TAAATTTAAA 4380
 I D L E L L D T E K S G K I K S H K F K
 ATTATAGGAA TCTTTTCTGG TAAAAACAG GAAACATATA CAGGATTATC ATCTGATTTT 4440
 I I G I F S G K K Q E T Y T G L S S D F
 AGCGAAAATA TGGTTTTTGT AGATTATCA ACTAGCCAAG AAATATTAAA TAAATCAGAG 4500
 S E N M V F V D Y S T S Q E I L N K S E
 AATAATAGAA TTGCAAATAA AATTTTAATG TATTCTGGTA GTTTAGAATC TACAGAGCTT 4560
 N N R I A N K I L M Y S G S L E S T E L
 GCCTTAAACA AATTGAAAGA CTTTAAAATT GATAAGTCAA AGTATTCTAT TAAGAAAGAT 4620

A L N K L K D F K I D K S K Y S I K K D
AATAAAGCAT TCGAAGAGTC TTTAGAGTCA GTGAGTGGAA TAAAACATAT AATTAAAATA 4680
N K A F E E S L E S V S G I K H I I K I
ATGACTTATT CGATTATGTT AGGTGGAATA GTTGTCTTT CATTAACTTT GATTCTATGG 4740
M T Y S I M L G G I V V L S L I L I L W
TTAAGAGAAA GAATTTATGA AATAGGTATA TTTTATCTA TTGGAACAAC TAAGATACAA 4800
L R E R I Y E I G I F L S I G T T K I Q
ATTATAAGGC AATTTATATT TGAGTTAATA TTCATATCAA TACCAAGTAT AATATCCTCC 4860
I I R Q F I F E L I F I S I P S I I S S
TTATTTTTAG GGAATCTACT ATTAAAAGTA ATTGTAGAAG GATTTATTAA CTCAGAGAAC 4920
L F L G N L L L K V I V E G F I N S E N
TCAATGATTT TCGGTGGAAG TTTAATAAAT AAAAGCAGTT TTATGTTAAA CATAACAACA 4980
S M I F G G S L I N K S S F M L N I T T
CTTGCGAGAAA GTTATTTAAT ATTAATAAGT ATTATTGTTT TATCAGTTGT AATGGCCTCT 5040
L A E S Y L I L I S I I V L S V V M A S
TCATTAATAT TATTTAAGAA ACCACAAGAA ATATTATCAA AAATAAGTTA GGAGCAAATA 5100
S L I L F K K P Q E I L S K I S .
ATGGATATAT TAGAAATAAA GAATGTAAAT TACAGTTACG CAAATTCTAA AGAAAAAGTT 5160
M D I L E I K N V N Y S Y A N S K E K V
|---->
TTGTCAGGAG TAAATCAAAA ATTTGAACTT GGAAAGTTTT ATGCGATAGT AGGGAAGTCA 5220
L S G V N Q K F E L G K F Y A I V G K S
GGAACAGGAA AATCCACACT TCTTTCCTTA CTTGCGAGGAC TTGATAAAGT TCAAACAGGA 5280
G T G K S T L L S L L A G L D K V Q T G
AAAATCTTGT TTAAGAATGA AGATATAGAA AAGAAAGGAT ATAGTAATCA CAGAAAAAAT 5340
K I L F K N E D I E K K G Y S N H R K N
AATATATCTT TGGTATTTCA AAATTATAAT TTAATAGATT ATTTATCGCC GATTGAAAAT 5400
N I S L V F Q N Y N L I D Y L S P I E N
ATTAGACTAG TAAATAAATC AGTAGATGAG AGTATCTTGT TCGAATTAGG TTTAGATAAA 5460
I R L V N K S V D E S I L F E L G L D K
AAACAAATAA AAAGAAATGT TATGAAATTA TCTGGTGGTC AGCAACAAAG GGTAGCTATT 5520
K Q I K R N V M K L S G G Q Q Q R V A I
GCTAGGGCAC TGGTATCAGA TGCCCCAATA ATRACTAGCTG ATGAGCCTAC CGGTAACCTA 5580
A R A L V S D A P I I L A D E P T G N L
GACAGTGTTA CTGCTGGAGA AATAATT (SEQ ID NO:27) 5607
D S V T A G E I I .

FIG. 5a

IQSLTEGQLR SDIPEFRAGD TVRVHAKVVE GTRERIQIFE GVVISRKGQG 50
 ISEMYTVRKI SGGIGVERTF PIHTPRVDKI EVVRYGKVRR AKLYYLRALQ 100
 GKAARIKEIR R (SEQ ID NO:28) 111

FIG. 5b

MRFAECLGLT VNDIDYTNKY LSINKTWDYH FNQRYLPTKN KSSIRNIPID 50
 NDTLFFLHEF TKNKNDRLED KLSNNAVNKT IRKITGREVR VHSLRHTFAS 100
 YLISISQVLD HENLNITLEV YAHQLQEOKD RNDKLNQRNL GQNSSKPLET 150
 CNEYVPCRNR TSNYSLGGSC YIH (SEQ ID NO:29) 173

FIG. 5c

MKSSNEIEKA LYESSNSSIS ITKKDGKYFN INQFKNIEKI KEVEEKIFQY 50
 DGLAKLKDLDK VVSGEQSINR EDLSDEFKNV VSLEATSNTK RNLLFSSGVF 100
 SFKEGKNIEE NDKNSILVHE EFAKQNKLLK GDEIDLELLD TEKSGKIKSH 150
 KFKIIGIFSG KKQETYTGLS SDFSENMFV DYSTSQEILN KSENNRIANK 200
 ILMYSGSLES TELALNKLKD FKIDKSKYSI KKDKNKAFES LESVSGIKHI 250
 IKIMTYSIML GGIVVLSLIL ILWLRERIYE IGIFLSIGTT KIQIIRQFIF 300
 ELIFISIPSI ISSLFLGNLL LKVIVEGFIN SENSMIFGGS LINKSSFMLN 350
 ITTLAESYLI LISIIVLSVV MASSLILFKK PQEILSKIS 389
 (SEQ ID NO:30)

FIG. 5d

MDILEIKNVN YSYANSKEKV LSGVNQKFEL GKFYAIVGKS GTGKSTLLSL 50
 LAGLDKVQTG KILFKNEDIE KKGYSNHRKN NISLVFQNYN LIDYLSPIEN 100
 IRLVNKSVDE SILFELGLDK KQIKRNVMKL SGGQQQRVAI ARALVSDAPI 150
 ILADEPTGNL DSVTAGEII (SEQ ID NO:31) 169

FIG. 5e

CATATGACAA	TATTTTTCAA	AGTCTACATC	ACTTACTCGC	CTGTCTGGA	AAATCTGGCA	60
ATACATTAAT	CGACCAATTA	GTTGCTGATG	GTTTACTTCA	TGCAGATAAT	CACTACCATT	120
TTTTCAATGG	GAAGTCTCTG	GCCACTTTCA	ATACTAACCA	ATTGATTTCG	GAAGTTGTCT	180
ATGTTGAAAT	ATCCTTAGAT	ACTATGTCTA	GTGGTGAACA	TGATTTAGTA	AAAGTTAACA	240
TTATCAGACC	CACTACCGAG	CATACTATCC	CCACGATGAT	GACAGCTAGC	CCCTATCATC	300
AAGGTATCAA	TGATCCTGCC	GCAGACCAA	AAACATACCA	AATGGAGGGT	GCGCTAGCAG	360
TTAAACAGCC	TAAACACATA	CAAGTTGACA	CAAACCATT	TAAAGAAGAA	GTAAAACATC	420
CTTCAAAATT	ACCCATCAGC	CCTGCAACTG	AAAGCTTCAC	ACACATTGAC	AGTTATAGTC	480
TCAATGACTA	TTTTCTTTCT	CGTGGTTTTG	CTAATATATA	CGTTTCAGGT	GTGGGTACTG	540
CTGGCTCTAC	GGGTTTCATG	ACCAGTGGGG	ATTACCAACA	AATACAAAGC	TTTAAAGCAG	600
TCATTGATTG	GTTAAATGGT	AAGGTTACTG	CATTACAAG	TCATAAACGA	GATAAACAAAG	660
TCAAGGCTGA	TTGGTCAAAC	GGCCTTGATG	CAACCACAGG	TAAATCTTAT	CTCGGTACCA	720
TGTCAACTGG	TTTAGCAACA	ACTGGCGTTG	AGGGGCTGAA	AGTCATTATC	GCTGAAGCCG	780
CAATCTCCAC	ATGGTATGAT	TATTATCGAG	AAAATGGGCT	TGTGTGTAGT	CCAGGCGGCT	840
ACCCCGGTGA	AGATTTAGAC	GTTTTAACAG	AATTAACATA	CTCACGAAAC	CTCTTAGCTG	900
GTGATTACAT	CAAAAACAAC	GATTGCTATC	AAGCATTGTT	AAATGAACAA	TCAAAAAGCAA	960
TTGACCGTCA	AAGTGGGGAT	TACAACCAAT	ACTGGCATGA	CCGTAATTAC	CTAACTCAG	1020
TCAATAATGT	CAAAGTCGA	GTAGTTTACA	CTCATGGACT	ACAGGATTGG	AATGTTAAGC	1080
CAAGACATGT	CTACAAAGTT	TTCAATGCAT	TGCCTCAAAC	CATCAAAAAA	CACCTTTTTT	1140
TACATCAAGG	TCAACATGTG	TATATGCATA	ATTGGCAGTC	GATTGATTTT	CGTGAAGCA	1200
TGAATGCCTT	ACTAAGCCAA	GAACTACTTG	GCATTGACAA	TCATTTCCAA	TTAGAAGAGG	1260
TCATTTGGCA	AGATAATACT	ACTGAGCAA	CTTGGCAAGT	TTAGATGCT	TTCGGAGGAA	1320
ACCATCAAGA	GCAAATTGGT	TTAGGTGATA	GTAAAAAAT	TATTGATAAC	CATTATGACA	1380
AAGAAGCCTT	TGATACTTAT	TGTAAAGACT	TCAATGTGTT	CAAAAATGAT	CTTTTCAAGG	1440
GAAATAATAA	AACCAATCAA	ATCACTATTA	ATCTTCCTCT	AAAGAAAAAT	TATCTCCTGA	1500
ATGGACAGTG	CAAACCTCCAT	CTACGTGTTA	AAACTAGTGA	CAAAAAGGCC	ATTTTATCAG	1560
CCCAAATCTT	AGACTATGGT	CCTAAAAAAC	GATTCAAAGA	TACACCAACC	ATCAAATTCT	1620
TAAACAGCCT	TGATAATGGT	AAAAATTTTG	CCAGAGAAGC	TTTACGTGAA	CTCCCGTTTA	1680
CTAAAGATCA	TTATCGTGTC	ATCAGTAAAG	GTGTCTTGAA	CCTTCAAAAT	CGTACAGACT	1740
TACTTACAAT	TGAGGCTATC	GAGCCAGAAC	AATGGTTTGA	TATCGAGTTT	AGCCTCCAAC	1800
CAAGTATATA	TCAATTGAGT	AAAGGTGATA	ATCTAAGGAT	TATCCTTTAT	ACAACCTGATT	1860
TTGAACATAC	CATTTCGAGAT	AATGCTAGTT	ACTCTATAAC	AGTAGATTTG	AGTCAATCTT	1920
ATTTAACTAT	CCCAACTAAT	CAAGGAAATT	AACTTATGAA	ACTTCTTACT	AAAGAACGGT	1980
TTGATGATTC	TCAACACTTT	TGGTACCAGA	TCAATTTATT	ACAAGAGAGT	AACTTCGGAG	2040
CAGTTTTTGA	CCATGATAAT	AAAAACATTC	CACAGGTTGT	TGCAACTATT	GTTGATGATT	2100
TACAAGGTTT	CGGAAGTTTC	AATCATTCTT	GGTATTTTGG	CAATACTACT	GATACTTCCA	2160
TCCTTATGAT	TGCTCATTTA	AATCGAAAAT	TCTATATTCA	GGTTAATTTA	AAGGACTTTG	2220
ACTTTGCACT	CAATTTAATA	GCTATAAATA	ATTGGAAGAG	TCTCCTCCAA	ACTCAACTTG	2280
AAGCTCTAAA	CGATACCCTA	GCAATATTTT	AATAAATAAG	GTAGAATGGA	GTGACAAAGC	2340
AACGCGAGGG	AGACTGATTA	ATGTCATCTT	ATTGGAATAA	CTATCCTGAA	CTTAAAAAAA	2400

ATATTGATGA	AACCAATCAA	CTAATTCAAG	AAAGAATACA	GGTCAGAAAT	AAAGATATTG	2460
AAGCGGCGCT	AAGCCAACTC	ACAGCTGCGG	GAGGAAAACA	GCTCAGACCA	GCATTCTTTT	2520
ACCTTTTTTC	TCAACTTGGT	AATAAGGAGA	ATCAAGATAC	TCAGCAACTA	AAGAAAATCG	2580
CTGCTTCTTT	AGAAATCCTT	CACGTTGCTA	CATTAATCCA	TGATGATGTC	ATTGATGACT	2640
CACCACTAAG	ACGTGGAAAT	ATGACCATTC	AAAGCAAGTT	TGGCAAAGAC	ATCGCAGTTT	2700
ATACTGGGGA	TTTACTTTTC	ACAGTCTTTT	TCGATCTTAT	TTTAGAATCT	ATGACTGATA	2760
CACCATTTAT	GAGGATTAAT	GCAAAAATCTA	TGCGTAAAAAT	TCTCATGGGA	GAATTGGACC	2820
AGATGCACCT	TCGTTACAAT	CAACAACAAG	GTATCCATCA	CTATTTACGT	GCGATTTTCAG	2880
GTAAGACAGC	CGAACTCTTT	AAATTAGCTA	GCAAAGAAGG	AGCTTACTTT	GGTGGTGCAG	2940
AGAAGGAGGT	TGTTCTGCTA	GCAGGCCATA	TCGGCTTTAA	CATTGGTATG	ACATTCCAAA	3000
TTTTGGATGA	TATCCTGGAT	TATACTGCAG	ATAAAAAAAC	ATTTAATAAG	CCTGTCTTAG	3060
AGGATTTAAC	ACAAGGCGTT	TACAGCCTTC	CTCTACTTCT	TGCCATTGAA	GAAAATCCTG	3120
ATATTTTCAA	ACCTATTTTA	GATAAAAAAA	CAGATATGGC	TACTGAAGAC	ATGGAAAAAA	3180
TTGCTTATCT	CGTCGTTTCC	CATAGAGGTG	TTGACAAAGC	TCGCCATCTA	GCTCGTAAAT	3240
TTACTGAGAA	AGCTATTAGT	GACATAAATA	AGCTACCCCA	GAAGTCTGCA	AAAAAACAGT	3300
TGCTACAATT	AACTAATTAC	CTTTTAAAAC	GCAAAATTTA	AATAATAAAA	AAACATTCCA	3360
CAATGCTAGA	AAAGCAGTTA	GGGAATGTTT	TTTTATTATC	ATTTATTTAT	CGCACCTATC	3420
AATCATCATA	GATCACCATC	ATCAGCGGCT	TTCAGCTGAC	GGTAACGTTG	ACTACTTTGA	3480
GACAATCTCT	GAGGAGAACC	TTCCAACCTC	AATTGCCCAT	TTTCTATAAA	TAAGATACGA	3540
TCAGCATGTT	CAATACCTTT	TAAGTGATGT	GTAATCCAAA	CTAAGGTCTT	ACCTTCCAAT	3600
TCTTTCATAA	ATACCCTTAG	TAAGGCTTGT	TCAGTAATAG	GATCAAGTCC	AACAGTTGGC	3660
TCATCTAAGA	TAACAATTGG	GACATCTTTT	AGTAAGATTG	TAGCCAAAGC	AATTCTATGC	3720
CTTTCGCCAC	CTGAAAACCT	AAGTCCAGCT	TCATCAACCA	TTGTATAGAG	ACCATCTGAT	3780
AAATCAGTGA	CCATCTCTTT	CAATCCAAC	CGTTCAAGAA	CTTTCATAC	ATCTTCTTCA	3840
CTAGCATCTT	GGTTTCCAAT	GCGAATGTTA	TTTAGCAGGG	TTGTATTAAA	AAGGTAGGGC	3900
GCTTGTGTA	TCACTCCAAT	ATAGTTAGAA	ATGCAATCAC	CAACTATTGA	AACATCAGCA	3960
CCGCCTAGGG	TAATCTTCCC	TTGACTTGCT	TTCAAGTCGC	CACGAAGTAG	ACTAGCTAAG	4020
GTACTCTTGC	CAGAACCACT	CCGCCCTAAA	ATAGCAATTT	TTTCTCCTTC	TTTAATATCC	4080
AAATCTAAAT	GATGCAAAAC	CCATTTCTCT	TGTGGCTTAT	ACTGGAAACT	TAAATTCTTG	4140
ACGGAAAAAT	CATATGGCTT	ATTAGGCAAT	T (SEQ ID NO:32)			4171

FIG. 6a

YDNIFQSLHH LLACRGKSGN TLIDQLVADG LLHADNHYHF FNGKSLATFN 50
 TNQLIREVVY VEISLDTMSS GEHDLVKVNI IRPTTEHTIP TMMTASPYHQ 100
 GINDPAADQK TYQMEGALAV KQPKHIQVDT KPFKEEVKHP SKLPISPATE 150
 SFTHIDSYSL NDYFLSRGFA NIYVSGVGTA GSTGFMTSGD YQQIQSEFKAV 200
 IDWLNKGKVT FTSHKRDKQV KADWSNGLVA TTGKSYLGTM STGLATTGVE 250
 GLKVIIAEEA ISTWYDYRE NGLVCSPGGY PGEDLDVLTE LTYSRNLLAG 300
 DYIKNNDCYQ ALLNEQSKAI DRQSGDYNQY WHDRNYLTHV NNVKSRVVYT 350
 HGLQDWNVKP RHVYKVFNAL PQTIKKHLFL HQGQHVYMHN WQSIDFRESM 400
 NALLSQELLG IDNHFQLEEV IWQDNTTEQT WQVLDAFGGN HQEQIGLGDS 450
 KKLIDNHYDK EAFD TYCKDF NVFKNDLFGK NNKTNQITIN LPLKKNYLLN 500
 GQCKLHLRVK TSDKKAILS A QILDYGPCKR FKDTPTIKFL NSLDNGKNFA 550
 REALREL PFT KDHYRVISKG VLN LQNRTDL LTIEAIEPEQ WFDIEFSLQP 600
 SIYQLSKGDN LRILYTTDF EHTIRDNASY SITVDLSQSY LTIPTNQGN 649

(SEQ ID NO:33)

FIG. 6b

MKLLTKERFD DSQHFYQIN LLQESNFGAV FDHDNKNIPQ VVATIVDDLQ 50
 GSGSSNHFWY FGNTTDT SIL MIAHLNRKFY IQVNLKDFDF ALNLIANNW 100
 KSL LQTQLEA LNDTLAIFQ (SEQ ID NO:34) 119

FIG. 6c

MSSYWNNYPE LKKNIDETNQ LIQERIQVRN KDIEAALS QL TAAGGKQLRP 50
 AFFYLFSQLG NKENQDTQQL KKIAASLEIL HVATLIHDDV IDDSPLRRGN 100
 MTIQSKFGKD IAVYTGDLLE TVFFDLILES MTDTPFMRIN AKSMRKILMG 150
 ELDQMHLRYN QQQGIHHYLR AISGKTAE LF KLASKEGAYF GGAEKEVVRL 200
 AGHIGFNIGM TFQILDDILD YTADKKT FNK PVLEDLTQGV YSLPLLLAIE 250
 ENPDIFKPIL DKKTDMATED MEKIAYLVVS HRGVDKARHL ARKFTEKAIS 300
 DINKLPQNSA KKQLLQLTNY LLKRKI (SEQ ID NO:35) 326

FIG. 6d

LPNKPYDFSV KNLSFQYKPQ EKWVLHHLDL DIKEGEKIAI LGRSGSGKST 50
LASLLRGDLK ASQGKITLGG ADVSIVGDCI SNYIGVIQQA PYLFNTLLN 100
NIRIGNQDAS EEDVWKVLER VGLKEMVTDL SDGLYTMVDE AGLRFSGGER 150
HRIALARILL KDVPIVILDE PTVGLDPITE QALLRVFMKE LEGKTLVWIT 200
HHLKGIEHAD RILFIENGQL ELEGSPQELS QSSQRYRQLK AADDGDL 247

(SEQ ID NO:36)

FIG. 6e

AATTCTATTT	GGAGGTTTTT	CTTGAATAAA	TGGTTAGTTA	AGGCAAGTTC	CTTAGTTGTT	60
TTAGGTGGTA	TGTTTTTATC	TGCGGGTTCC	CGAGTTTTAG	CGGATACTTA	TGTCCGTCCA	120
ATTGATAATG	GTAGAATTAC	AACAGGTTTC	AATGGTTATC	CTGGACATTG	TGGGGTGGAT	180
TATGCTGTTC	CGACTGGAAC	GATTATTAGG	GCAGTGGCAG	ATGGTACTGT	GAAATTTGCA	240
GGAGCTGGAG	CCAACTTTTC	TTGGATGACA	GACTTAGCAG	GAAATTGTGT	CATGATTCAA	300
CATGCGGATG	GAATGCATAG	TGGTTACGCT	CATATGTCAC	GTGTGGTGGC	TAGGACTGGG	360
GAAAAAGTCA	AACAAGGAGA	TATCATCGGT	TACGTAGGAG	CAACTGGTAT	GGCGACGGGA	420
CCTCACCTTC	ATTTTGAATT	TTTACCAGCT	AACCCTAATT	TTCAAAATGG	TTTCCATGGA	480
CGTATCAATC	CAACGTCACT	AATTGCTAAC	GTTGCGACCT	TTAGTGGAAA	AACGCAAGCA	540
TCAGCTCCAA	GCATTAAGCC	ATTACAATCA	GCTCCTGTAC	AGAATCAATC	TAGTAAATTA	600
AAAGTGTATC	GAGTAGATGA	ATTACAAAAG	GTTAATGGTG	TTTGGTTAGT	CAAAAATAAC	660
ACCTAACGC	CGACTGGGTT	TGATTGGAAC	GATAATGGTA	TACCAGCATC	AGAAATTGAT	720
GAGGTTGATG	CTAATGGTAA	TTTGACAGCT	GACCAGGTTT	TTCAAAAAGG	TGGTTACTTT	780
ATCTTTAATC	CTAAAACCTC	TAAGACTGTA	GAAAACCCA	TCCAAGGAAC	AGCTGGTTTA	840
ACTTGGGCTA	AGACACGCTT	TGCTAATGGT	AGTTCAGTTT	GGCTTCGCGT	TGACAACAGT	900
CAAGAACTGC	TTTACAAATA	GTTTGAGGTA	TTGATTCATT	GTTTTAAATG	ACAGTTTTGT	960
TACTAACTAA	GTACAATTTT	TTTAAACCGT	CTGAAAATAA	TTTTTATAGT	CAGTAAAGTG	1020
TGATATTATA	GTCTCGGACT	AATAAAAAGG	AAATAGGAAT	TGAAGCAATG	AAAATGAATA	1080
AAAAGGTAAT	ATTGACATCG	ACAATGGCAG	CTTCGCTATT	ATCAGTCGCA	AGTGTTC AAG	1140
CACAAGAAAC	AGATACGACG	TGGACAGCAC	GTACTGTTTC	AGAGGTAAG	GCTGATTTGG	1200
TAAAGCAAGA	CAATAAATCA	TCATATACTG	TGAAATATGG	TGATACTACTA	AGCGTTATTT	1260
CAGAAGCAAT	GTCAATTGAT	ATGAATGTCT	TAGCAAAAAT	TAATAACATT	GCAGATATCA	1320
ATCTTATTTA	TCCTGAGACA	ACACTGACAG	TAACTTACGA	TCAGAAGAGT	CATACTGCCA	1380
CTTCAATGAA	AATAGAAACA	CCAGCAACAA	ATGCTGCTGG	TCAAACAACA	GCTACTGTGG	1440
ATTTGAAAAC	CAATCAAGTT	TCTGTTGCAG	ACCAAAAAGT	TTCTCTCAAT	ACAATTTCCG	1500
AAGGTATGAC	ACCAGAAGCA	GCAACAACGA	TTGTTTCGCC	AATGAAGACA	TATTCCTCTG	1560
CGCCAGCTTT	GAAATCAAAA	GAAGTATTAG	CACAAGAGCA	AGCTGTTAGT	CAAGCAGCAG	1620
CTAATGAACA	GGTATCAACA	GCTCCTGTGA	AGTCGATTAC	TTCAGAAGTT	CCAGCAGCTA	1680
AAGAGGAAGT	TAAACCAACT	CAGACGTCAG	TCAGTCAGTC	AACAACAGTA	TCACCAGCTT	1740
CTGTTGCCGC	TGAAACACCA	GCTCCAGTAG	CTAAAGTAGC	ACCGGTAAGA	ACTGTAGCAG	1800
CCCCTAGAGT	GGCAAGTGTT	AAAGTAGTCA	CTCCTAAAGT	AGAAACTGGT	GCATCACCAG	1860
AGCATGTATC	AGCTCCAGCA	GTTCCCTGTGA	CTACGACTTC	AACAGCTACA	GACAGTAAAGT	1920
TACAAGCGAC	TGAAGTTAAG	AGCGTTCGG	TAGCACAAAA	AGCTCCAACA	GCAACACCGG	1980
TAGCACAACC	AGCTTCAACA	ACAAATGCAG	TAGCTGCACA	TCCTGAAAAT	GCAGGGCTCC	2040
AACCTCATGT	TGCAGCTTAT	AAAGAAAAAG	TAGCGTCAAC	TTATGGAGTT	AATGAATTCA	2100
GTACATACCG	TGCAGGTGAT	CCAGGTGATC	ATGGTAAAGG	TTTAGCAGTC	GACTTTATTG	2160
TAGGTAAAAA	CCAAGCACTT	GGTAATGAAG	TTGCACAGTA	CTCTACACAA	AATATGGCAG	2220
CAAATAACAT	TTCATATGTT	ATCTGGCAAC	AAAAGTTTTA	CTCAAATACA	AATAGTATTT	2280
ATGGACCTGC	TAATACTTGG	AATGCAATGC	CAGATCGTGG	TGGCGTTACT	GCCAACCATT	2340
ATGACCATGT	TCACGTATCA	TTTAACAAAAT	AATATAAAAA	AGGAAGCTAT	TTGGCTTCTT	2400

TTTTATATGC CTTGAATAGA CTTTCAAGGT TCTTATCTAA TTTTATTAA ATTGAGGAGA 2460
 TTAAGCTATA AGTCTGAAAC TACTTTCACG TTAACCGTGA CTAAATCAA ACGTTAAAAC 2520
 TAAAATCTAA GTCTGTAAAG ATTATTGAAA ACGCTTTAAA AACAGATATA ATAAGGTTTG 2580
 TAGATATCTA AAATTAATAA AGATAAGGAA GTGAGAATAT GCCACATCTA AGTAAAGAAG 2640
 CTTTTAAAAA GCAAATAAAA AATGGCATT TGTGTTCATG TCAAGCTTTG CCTGGGGAGC 2700
 CTCTTTTATAC TGAAAGTGGA GGTGTTATGC CTCTTTTAGC TTTGGCAGCT CAAGAAGCAG 2760
 GAGCGGTTGG TATAAGAGCC AATAGTGTCC GCGACATTAA GGAAATCAA GAAGTTACTA 2820
 ATTTACCTAT CATCGGCATT ATTAACCGTG AATATCCTCC ACAAGAACCA TTTATCACTG 2880
 CTACGATGAC AGAGGTGGAT CAATTAGCTA GTTTAGATAT TGCAGTAATA GCCTTAGATT 2940
 GTACACTTAG AGAGCGTCAT GATGGTTTGA GTGTAGCTGA GTTTATTCAA AAGATAAAAG 3000
 GGAAATATCC TGAACAGTTG CTAATGGCTG ATATAAGTAC TTTTGAAGAA GGTAAAAATG 3060
 CTTTTGAAGC AGGAGTTGAT TTTGTGGGTA CAACTCTATC TGGATACACA GATTACAGCC 3120
 GCCAAGAAGA AGGACCGGAT ATAGAACTCC TTAATAAGCT TTGTCAAGCC GGTATAGATG 3180
 TGATTGCGGA AGGTAATAAT CATACTCCTA AGCAAGCTAA TGAAATTAAT CATATAGGTG 3240
 TTGCAGGAAT TGTAGTTGGT GGTGCTATCA CTAGACCAA AGAAATAGCG GAGCGTTTCA 3300
 TCTCAGGACT TAGTAAAAG TGTTACTCAA AAATCAAAT CAAAATAAAA AAGGGGAATA 3360
 GTTATGAGTA TCAAAAAAAG TGTGATTGGT TTTGCCTCG GAGCTGCAGC ATTATCAATG 3420
 TTTGCTTGTG TAGACAGTAG TCAATCTGTT ATGGCTGCCG AGAAGGATAA AGTCGAAATT 3480
 (SEQ ID NO:37)

FIG. 7a

NSIWRFFLNK WLVKASSLVV LGGMVL SAGS RVLADTYVRP IDNGRITTGF 50
 NGYPGHCGVD YAVPTGTIIR AVADGTVKFA GAGANFSWMT DLAGNCVMIQ 100
 HADGMHSGYA HMSRVVARTG EKVKQGDII G YVGATGMATG PHLHFEFLPA 150
 NPNFQNGFHG RINPTSLIAN VATFSGKTQA SAPSIKPLQS APVQNQSSKL 200
 KVYRVDELQK VNGVWLKNN TLTPTGFDWN DNGIPASEID EVDANGNLTA 250
 DQVLQKGGYF IFNPKTLKTV EKPIQGTAGL TWAKTRFANG SSVWLRVDNS 300
 QELLYK (SEQ ID NO:38) 306

FIG. 7b

MKMNKKVLLT	STMAASLLSV	ASVQAQETDT	TWTARTVSEV	KADLVKQDNK	50
SSYTVKYGDT	LSVISEAMSI	DMNVLAKINN	IADINLIYPE	TTLTVTYDQK	100
SHTATSMKIE	TPATNAAGQT	TATVDLKTNO	VSVADQKVSL	NTISEGMTPE	150
AATTIVSPMK	TYSSAPALKS	KEVLAQEQAV	SQAAANEQVS	TAPVKSITSE	200
VPAAKEEVKP	TQTSVSQSTT	VSPASVAAET	PAPVAKVAPV	RTVAAPRVAS	250
VKVVTPKVET	GASPEHVSAP	AVPVTTTSTA	TDSKLQATEV	KSVPVAQKAP	300
TATPVAQPAS	TTNAVAHPE	NAGLQPHVAA	YKEKVASTYG	VNEFSTYRAG	350
DPGDHGKGLA	VDFIVGKNQA	LGNEVAQYST	QNMAANNISY	VIWQQKFYSN	400
TNSIYGPANT	WNAMPDRGGV	TANHYDHVHV	SFNK	(SEQ ID NO:39)	434

FIG. 7c

MPHLSKEAFK	KQIKNGIIVS	COALPGEPLY	TESGGVMPLL	ALAAQEAGAV	50
GIRANSVRDI	KEIQEVTNLP	IIGIIKREYP	PQEPFITATM	TEVDQLASLD	100
IAVIALDCTL	RERHDGLSVA	EFIQKIKGKY	PEQLLMADIS	TFEEGKNAFE	150
AGVDFVGTTL	SGYTDYXRQE	EGPDIELLNK	LCQAGIDVIA	EGKIHTPKQA	200
NEINHIGVAG	IVVGGAITRP	KEIAERFISG	LS	(SEQ ID NO:40)	232

FIG. 7d

MSIKKSVIGF	CLGAAALSMF	ACVDSSQSVM	AAEKDKVEI	39
(SEQ ID NO:41)				

FIG. 7e

ATGAAAATGA	ATAAAAAGGT	ACTATTGACA	TCGACAATGG	CAGCTTCGCT	50
ATTATCAGTC	GCAAGTG TTC	AAGCACAAGA	AACAGATACG	ACGTGGACAG	100
CACGTACTGT	TTCAGAGGTA	AAGGCTGATT	TGGTAAAGCA	AGACAATAAA	150
TCATCATATA	CTGTGAAATA	TGGTGATACA	CTAAGCGTTA	TTTCAGAAGC	200
AATGTCAATT	GATATGAATG	TCTTAGCAAA	AATTAATAAC	ATTGCAGATA	250
TCAATCTTAT	TTATCCTGAG	ACAACACTGA	CAGTAACTTA	CGATCAGAAG	300
AGTCATACTG	CCACTTCAAT	GAAAATAGAA	ACACCAGCAA	CAAATGCTGC	350
TGGTCAAACA	ACAGCTACTG	TGGATTTGAA	AACCAATCAA	GTTTCTGTTG	400
CAGACCAAAA	AGTTTCTCTC	AATACAATTT	CGGAAGGTAT	GACACCAGAA	450
GCAGCAACAA	CGATTGTTTC	GCCAATGAAG	ACATATTCTT	CTGCGCCAGC	500
TTTGAAATCA	AAAGAAGTAT	TAGCACAAGA	GCAAGCTGTT	AGTCAAGCAG	550
CAGCTAATGA	ACAGGTATCA	ACAGTCTCTG	TGAAGTCGAT	TACTTCAGAA	600
GTTCCAGCAG	CTAAAGAGGA	AGTTAAACCA	ACTCAGACGT	CAGTCAGTCA	650
GTCAACAACA	GTATCACCAG	CTTCTGTTGC	CGTGAAACA	CCAGTCCAG	700
TAGCTAAAGT	AGCACCGGTA	AGAACTGTAG	CAGCCCCTAG	AGTGGCAAGT	750
GTTAAAGTAG	TCACTCCTAA	AGTAGAAACT	GGTGCATCAC	CAGAGCATGT	800
ATCAGCTCCA	GCAGTTCCTG	TGACTACGAC	TTCAACAGCT	ACAGACAGTA	850
AGTTACAAGC	GACTGAAGTT	AAGAGCGTTC	CGGTAGCACA	AAAAGCTCCA	900
ACAGCAACAC	CGGTAGCACA	ACCAGCTTCA	ACAACAAATG	CAGTAGCTGC	950
ACATCCTGAA	AATGCAGGGC	TCCAACCTCA	TGTTGCAGCT	TATAAAGAAA	1000
AAGTAGCGTC	AACTTATGGA	GTTAATGAAT	TCAGTACATA	CCGTGCAGGT	1050
GATCCAGGTG	ATCATGGTAA	AGGTTTAGCA	GTCGACTTTA	TTGTAGGTAA	1100
AAACCAAGCA	CTTGGAATG	AAGTTGCACA	GTA CTCTACA	CAAAATATGG	1150
CAGCAAATAA	CATTTCATAT	GTTATCTGGC	AACAAAAGTT	TTACTCAAAT	1200
ACAAATAGTA	TTTATGGACC	TGCTAATACT	TGGAATGCAA	TGCCAGATCG	1250
TGGTGGCGTT	ACTGCCAACC	ATTATGACCA	TGTTACGTA	TCATTTAACA	1300
AATAA					1305

(SEQ ID NO:42)

FIG. 8

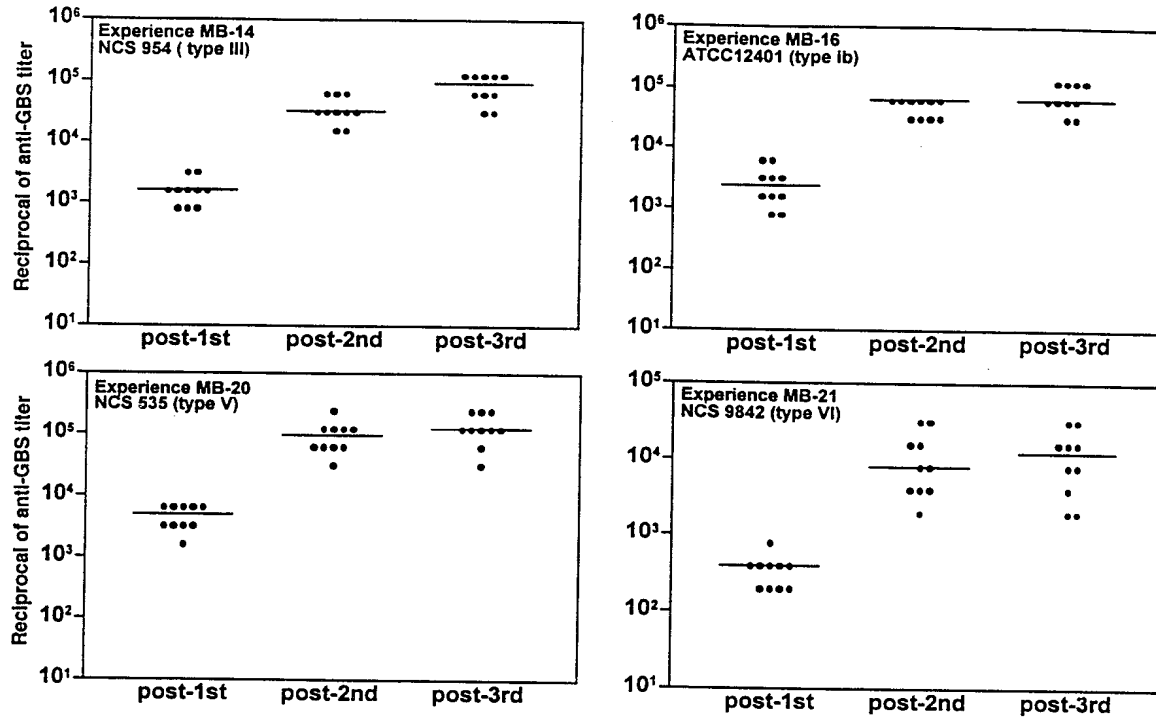
CAAGAAACAG	ATACGACGTG	GACAGCACGT	ACTGTTTCAG	AGGTAAAGGC	50
TGATTTGGTA	AAGCAAGACA	ATAAATCATC	ATATACTGTG	AAATATGGTG	100
ATACACTAAG	CGTTATTTCA	GAAGCAATGT	CAATTGATAT	GAATGTCTTA	150
GCAAAAATTA	ATAACATTGC	AGATATCAAT	CTTATTTATC	CTGAGACAAC	200
ACTGACAGTA	ACTTACGATC	AGAAGAGTCA	TACTGCCACT	TCAATGAAAA	250
TAGAAACACC	AGCAACAAAT	GCTGCTGGTC	AAACAACAGC	TACTGTGGAT	300
TTGAAAACCA	ATCAAGTTTC	TGTTGCAGAC	CAAAAAGTTT	CTCTCAATAC	350
AATTTTCGGAA	GGTATGACAC	CAGAAGCAGC	AACAACGATT	GTTTTCGCCAA	400
TGAAGACATA	TTCTTCTGCG	CCAGCTTTGA	AATCAAAAGA	AGTATTAGCA	450
CAAGAGCAAG	CTGTTAGTCA	AGCAGCAGCT	AATGAACAGG	TATCAACAGC	500
TCCTGTGAAG	TCGATTACTT	CAGAAGTTCC	AGCAGCTAAA	GAGGAAGTTA	550
AACCAACTCA	GACGTCAGTC	AGTCAGTCAA	CAACAGTATC	ACCAGCTTCT	600
GTTGCCCGCTG	AAACACCAGC	TCCAGTAGCT	AAAGTAGCAC	CGGTAAGAAC	650
TGTAGCAGCC	CCTAGAGTGG	CAAGTGTAA	AGTAGTCACT	CCTAAAGTAG	700
AAACTGGTGC	ATCACCAGAG	CATGTATCAG	CTCCAGCAGT	TCCTGTGACT	750
ACGACTTCAA	CAGCTACAGA	CAGTAAGTTA	CAAGCGACTG	AAGTTAAGAG	800
CGTTCGGGTA	GCACAAAAAG	CTCCAACAGC	AACACCGGTA	GCACAACCAG	850
CTTCAACAAC	AAATGCAGTA	GCTGCACATC	CTGAAAATGC	AGGGCTCCAA	900
CCTCATGTTG	CAGCTTATAA	AGAAAAAGTA	GCGTCAACTT	ATGGAGTTAA	950
TGAATTCAGT	ACATACCGTG	CAGGTGATCC	AGGTGATCAT	GGTAAAGGTT	1000
TAGCAGTCGA	CTTTATTGTA	GGTAAAAACC	AAGCACTTGG	TAATGAAGTT	1050
GCACAGTACT	CTACACAAAA	TATGGCAGCA	AATAACATTT	CATATGTTAT	1100
CTGGCAACAA	AAGTTTTTACT	CAAATACAAA	TAGTATTTAT	GGACCTGCTA	1150
ATACTTGGA	TGCAATGCCA	GATCGTGGTG	GCGTTACTGC	CAACCATTAT	1200
GACCATGTTC	ACGTATCATT	TAACAAATAA	(SEQ ID NO:43)		1230

FIG. 9

QETDFTWTAR	TVSEVKADLV	KQDNKSSYTV	KYGDTLVIS	EAMSIDMNVL	50
AKINNIADIN	LIYPETTLTV	TYDQKSHTAT	SMKIETPATN	AAGQTTATVD	100
LKTNQVSVAD	QKVS LNTISE	GMTPEAATTI	VSPMKTYSSA	PALKSKEVLA	150
QEQAVSQAAA	NEQVSTAPVK	SITSEVPAAK	EEVKPTQTSV	SQSTTVSPAS	200
VAAETPAPVA	KVAPVRTVAA	PRVASVKVVT	PKVETGASPE	HVSAPAVPVT	250
TTSTATDSKL	QATEVKSVPV	AQKAPTATPV	AQPASTTNAV	AAHPENAGLQ	300
PHVAAYKEKV	ASTYGVNEFS	TYRAGDPGDH	GKGLAVDFIV	GKNQALGNEV	350
AQYSTQNMMA	NNISYVIWQQ	KFYSNTNSIY	GPANTWNAMP	DRGGVTANHY	400
DHVHVSEFNK	(SEQ ID NO:44)				409

FIG. 9a

Fig. 10



SEQUENCE LISTING

<110> BioChem Vaccins
 RIOUX, Clément
 DENIS, Martin
 BRODEUR, Bernard R.
 HAMEL, Josée
 CHARLEBOIS, Isabelle
 BOYER, Martine

<120> NOVEL GROUP B STREPTOCOCCUS ANTIGENS

<130> 12806-9PCT

<150> 60/075,425

<151> 1998-02-20

<160> 44

<170> FastSEQ for Windows Version 3.0

<210> 1

<211> 4514

<212> DNA

<213> Streptococcus

<220>

<221> CDS

<222> (3)...(464)

<221> CDS

<222> (534)...(887)

<223>

<221> CDS

<222> (1024)...(1767)

<221> CDS

<222> (1841)...(4288)

<221> CDS

<222> (2735)...(4288)

<400> 1

ta tct ggc aaa gag cca gct aat cgt ttt agt tgg gct aaa aat aaa
 Ser Gly Lys Glu Pro Ala Asn Arg Phe Ser Trp Ala Lys Asn Lys
 1 5 10 15

2 / 63

tta tta atc aat gga ttc att gca act cta gca gca act atc tta ttt	95
Leu Leu Ile Asn Gly Phe Ile Ala Thr Leu Ala Ala Thr Ile Leu Phe	
20 25 30	
ttt gca gtt caa ttc ata ggt ctt aaa cca gat tac cct gga aaa acc	143
Phe Ala Val Gln Phe Ile Gly Leu Lys Pro Asp Tyr Pro Gly Lys Thr	
35 40 45	
tac ttt att atc cta ttg aca gca tgg act ttg atg gca tta gta act	191
Tyr Phe Ile Ile Leu Leu Thr Ala Trp Thr Leu Met Ala Leu Val Thr	
50 55 60	
gct tta gtg gga tgg gat aat agg tat ggt tcc ttc ttg tcg tta tta	239
Ala Leu Val Gly Trp Asp Asn Arg Tyr Gly Ser Phe Leu Ser Leu Leu	
65 70 75	
ata tta tta ttc cag ctt ggt tca agc gca gga act tac cca ata gaa	287
Ile Leu Leu Phe Gln Leu Gly Ser Ser Ala Gly Thr Tyr Pro Ile Glu	
80 85 90 95	
ttg agt cct aag ttc ttt caa aca att caa cca ttt tta ccg atg act	335
Leu Ser Pro Lys Phe Phe Gln Thr Ile Gln Pro Phe Leu Pro Met Thr	
100 105 110	
tac tct gtt tca gga tta aga gag acc atc tcg ttg acg gga gac gtt	383
Tyr Ser Val Ser Gly Leu Arg Glu Thr Ile Ser Leu Thr Gly Asp Val	
115 120 125	
aac cat caa tgg aga atg cta gta atc ttt tta gta tca tcg atg ata	431
Asn His Gln Trp Arg Met Leu Val Ile Phe Leu Val Ser Ser Met Ile	
130 135 140	
ctt gct ctt ctt att tat cgt aaa caa gaa gat taatagaaag tatctagtga	484
Leu Ala Leu Leu Ile Tyr Arg Lys Gln Glu Asp	
145 150	
tagactaaca gtatgatatg gtatgtcaaa gtatttagga ggagaagat atg tct act	542
Met Ser Thr	
155	
tta aca ata att att gca aca tta act gct ttg gaa cat ttt tat att	590
Leu Thr Ile Ile Ile Ala Thr Leu Thr Ala Leu Glu His Phe Tyr Ile	
160 165 170	
atg tat ttg gag acg tta gcc acc cag tca aat atg act ggg aag att	638
Met Tyr Leu Glu Thr Leu Ala Thr Gln Ser Asn Met Thr Gly Lys Ile	
175 180 185	
ttt agt atg tct aaa gaa gag ttg tca tat tta ccc gtt att aaa ctt	686
Phe Ser Met Ser Lys Glu Glu Leu Ser Tyr Leu Pro Val Ile Lys Leu	
190 195 200 205	
ttt aag aat caa ggt gta tac aac ggc ttg att ggc cta ttc ctc ctt	734
Phe Lys Asn Gln Gly Val Tyr Asn Gly Leu Ile Gly Leu Phe Leu Leu	
210 215 220	

tat ggg tta tat att tca cag aat caa gaa att gta gct gtt ttt tta	782
Tyr Gly Leu Tyr Ile Ser Gln Asn Gln Glu Ile Val Ala Val Phe Leu	
225 230 235	
atc aat gta ttg cta gtt gct att tat ggt gct ttg aca gtt gat aaa	830
Ile Asn Val Leu Leu Val Ala Ile Tyr Gly Ala Leu Thr Val Asp Lys	
240 245 250	
aaa atc tta tta aaa cag ggt ggt tta cct ata tta gct ctt tta aca	878
Lys Ile Leu Leu Lys Gln Gly Leu Pro Ile Leu Ala Leu Leu Thr	
255 260 265	
ttc tta ttt taatactact tagccgttcg atttagttga acggctttta	927
Phe Leu Phe	
270	
gtaatcattt ttttctcata atacaggtag tttaagtaat ttgtctttaa aaatagtata	987
atataactac gaattcaaag agaggtgact ttgatt atg act gag aac tgg tta	1041
Met Thr Glu Asn Trp Leu	
275	
cat act aaa gat ggt tca gat att tat tat cgt gtc gtt ggt caa ggt	1089
His Thr Lys Asp Gly Ser Asp Ile Tyr Tyr Arg Val Val Gly Gln Gly	
280 285 290	
caa ccg att gtt ttt tta cat ggc aat agc tta agt agt cgc tat ttt	1137
Gln Pro Ile Val Phe Leu His Gly Asn Ser Leu Ser Ser Arg Tyr Phe	
295 300 305 310	
gat aag caa ata gca tat ttt tct aag tat tac caa gtt att gtt atg	1185
Asp Lys Gln Ile Ala Tyr Phe Ser Lys Tyr Tyr Gln Val Ile Val Met	
315 320 325	
gat agt aga ggg cat ggc aaa agt cat gca aag cta aat acc att agt	1233
Asp Ser Arg Gly His Gly Lys Ser His Ala Lys Leu Asn Thr Ile Ser	
330 335 340	
ttc agg caa ata gca gtt gac tta aag gat atc tta gtt cat tta gag	1281
Phe Arg Gln Ile Ala Val Asp Leu Lys Asp Ile Leu Val His Leu Glu	
345 350 355	
att gat aaa gtt ata ttg gta ggc cat agc gat ggt gcc aat tta gct	1329
Ile Asp Lys Val Ile Leu Val Gly His Ser Asp Gly Ala Asn Leu Ala	
360 365 370	
tta gtt ttt caa acg atg ttt cca ggt atg gtt aga ggg ctt ttg ctt	1377
Leu Val Phe Gln Thr Met Phe Pro Gly Met Val Arg Gly Leu Leu Leu	
375 380 385 390	
aat tca ggg aac ctg act att cat ggt cag cga tgg tgg gat att ctt	1425
Asn Ser Gly Asn Leu Thr Ile His Gly Gln Arg Trp Trp Asp Ile Leu	
395 400 405	
tta gta agg att gcc tat aaa ttc ctt cac tat tta ggg aaa ctc ttt	1473
Leu Val Arg Ile Ala Tyr Lys Phe Leu His Tyr Leu Gly Lys Leu Phe	
410 415 420	

ccg tat atg agg caa aaa gct caa gtt att tcg ctt atg ttg gag gat	1521
Pro Tyr Met Arg Gln Lys Ala Gln Val Ile Ser Leu Met Leu Glu Asp	
425 430 435	
ttg aag att agt cca gct gat tta cag cat gtg tca act cct gta atg	1569
Leu Lys Ile Ser Pro Ala Asp Leu Gln His Val Ser Thr Pro Val Met	
440 445 450	
gtt ttg gtt gga aat aag gac ata att aag tta aat cat tct aag aaa	1617
Val Leu Val Gly Asn Lys Asp Ile Ile Lys Leu Asn His Ser Lys Lys	
455 460 465 470	
ctt gct tct tat ttt cca agg ggg gag ttt tat tct tta gtt ggc ttt	1665
Leu Ala Ser Tyr Phe Pro Arg Gly Glu Phe Tyr Ser Leu Val Gly Phe	
475 480 485	
ggg cat cac att att aag caa gat tcc cat gtt ttt aat att att gca	1713
Gly His His Ile Ile Lys Gln Asp Ser His Val Phe Asn Ile Ile Ala	
490 495 500	
aaa aag ttt atc aac gat acg ttg aaa gga gaa att gtt gaa aaa gct	1761
Lys Lys Phe Ile Asn Asp Thr Leu Lys Gly Glu Ile Val Glu Lys Ala	
505 510 515	
aat tga aaaagtcaaa tcaactgactt ctgtgattaa aattgtattt tttatatctg	1817
Asn *	
ttttagtgct tattattggt gaa atg att cat ttg aaa cga act att tct gtt	1870
Met Ile His Leu Lys Arg Thr Ile Ser Val	
520 525	
gag caa cta aag agt gtt ttt ggg caa tta tct cca atg aat ctt ttc	1918
Glu Gln Leu Lys Ser Val Phe Gly Gln Leu Ser Pro Met Asn Leu Phe	
530 535 540 545	
tta att atc ctt gtg ggg gtt atc gct gtc tta ccg aca acc gga tat	1966
Leu Ile Ile Leu Val Gly Val Ile Ala Val Leu Pro Thr Thr Gly Tyr	
550 555 560	
gac ttt gta ctg aat gga ctt tta cgt aca gat aaa agc aaa agg tat	2014
Asp Phe Val Leu Asn Gly Leu Leu Arg Thr Asp Lys Ser Lys Arg Tyr	
565 570 575	
att tta cag act agt tgg tgt atc aac act ttt aat aac ttg tca gga	2062
Ile Leu Gln Thr Ser Trp Cys Ile Asn Thr Phe Asn Asn Leu Ser Gly	
580 585 590	
ttc ggt ggc tta atc gat att ggg ttg cgc atg gct ttt tat ggt aaa	2110
Phe Gly Gly Leu Ile Asp Ile Gly Leu Arg Met Ala Phe Tyr Gly Lys	
595 600 605	
aaa ggt caa gag aag agt gac cta aga gaa gtg act cgt ttt tta ccc	2158
Lys Gly Gln Glu Lys Ser Asp Leu Arg Glu Val Thr Arg Phe Leu Pro	
610 615 620 625	

tat	ctt	att	tct	ggt	ctg	tca	ttt	att	agt	gtg	att	gcc	tta	atc	atg	2206
Tyr	Leu	Ile	Ser	Gly	Leu	Ser	Phe	Ile	Ser	Val	Ile	Ala	Leu	Ile	Met	
				630					635					640		
agc	cat	att	ttt	cat	gcc	aaa	gct	agt	ggt	gat	tac	tat	tat	ttg	gta	2254
Ser	His	Ile	Phe	His	Ala	Lys	Ala	Ser	Val	Asp	Tyr	Tyr	Tyr	Leu	Val	
			645					650						655		
tta	att	ggt	gct	agt	atg	tat	ttt	cct	ggt	att	tat	tgg	att	tct	ggt	2302
Leu	Ile	Gly	Ala	Ser	Met	Tyr	Phe	Pro	Val	Ile	Tyr	Trp	Ile	Ser	Gly	
		660					665					670				
cat	aaa	gga	agc	cat	tat	ttc	gga	gat	atg	cca	tct	agt	act	cgt	ata	2350
His	Lys	Gly	Ser	His	Tyr	Phe	Gly	Asp	Met	Pro	Ser	Ser	Thr	Arg	Ile	
	675					680					685					
aaa	tta	ggt	ggt	ggt	tct	ttt	ttt	gaa	tgg	gga	tgt	gcg	gcc	gca	gca	2398
Lys	Leu	Gly	Val	Val	Ser	Phe	Phe	Glu	Trp	Gly	Cys	Ala	Ala	Ala	Ala	
690					695					700					705	
ttt	ata	att	atc	ggt	tat	tta	atg	ggc	att	cat	cta	cca	ggt	tat	aaa	2446
Phe	Ile	Ile	Ile	Gly	Tyr	Leu	Met	Gly	Ile	His	Leu	Pro	Val	Tyr	Lys	
				710					715					720		
att	tta	cca	cta	ttt	tgt	att	ggt	tgt	gcc	gtc	ggg	att	gta	tcc	ctt	2494
Ile	Leu	Pro	Leu	Phe	Cys	Ile	Gly	Cys	Ala	Val	Gly	Ile	Val	Ser	Leu	
			725					730					735			
att	ccc	ggt	gga	tta	gga	agt	ttt	gaa	tta	ggt	cta	ttt	aca	ggg	ttt	2542
Ile	Pro	Gly	Gly	Leu	Gly	Ser	Phe	Glu	Leu	Val	Leu	Phe	Thr	Gly	Phe	
		740					745						750			
gct	gcc	gag	gga	cta	cct	aaa	gaa	act	gtg	ggt	gca	tgg	tta	tta	ctt	2590
Ala	Ala	Glu	Gly	Leu	Pro	Lys	Glu	Thr	Val	Val	Ala	Trp	Leu	Leu	Leu	
	755					760					765					
tat	cgt	tta	gcc	tac	tat	att	att	cca	ttc	ttt	gca	ggt	atc	tat	ttc	2638
Tyr	Arg	Leu	Ala	Tyr	Tyr	Ile	Ile	Pro	Phe	Phe	Ala	Gly	Ile	Tyr	Phe	
770					775					780					785	
ttt	atc	cat	tat	tta	ggt	agt	caa	ata	aat	caa	cgt	tat	gaa	aat	gtc	2686
Phe	Ile	His	Tyr	Leu	Gly	Ser	Gln	Ile	Asn	Gln	Arg	Tyr	Glu	Asn	Val	
				790					795					800		
ccg	aaa	gag	tta	gta	tca	act	ggt	cta	caa	acc	atg	gtg	agc	cat	ttg	2734
Pro	Lys	Glu	Leu	Val	Ser	Thr	Val	Leu	Gln	Thr	Met	Val	Ser	His	Leu	
			805					810								
atg	cgt	att	tta	ggt	gca	ttc	tta	ata	ttt	tca	aca	gca	ttt	ttt	gaa	2782
Met	Arg	Ile	Leu	Gly	Ala	Phe	Leu	Ile	Phe	Ser	Thr	Ala	Phe	Phe	Glu	
		820					825					830				
aat	att	act	tat	att	atg	tgg	ttg	cag	aag	cta	ggc	ttg	gac	cca	tta	2830
Asn	Ile	Thr	Tyr	Ile	Met	Trp	Leu	Gln	Lys	Leu	Gly	Leu	Asp	Pro	Leu	
	835					840					845					

6 / 63

caa gaa caa atg tta tgg cag ttt cca ggt tta ttg ctg ggg gtt tgt	2878
Gln Glu Gln Met Leu Trp Gln Phe Pro Gly Leu Leu Leu Gly Val Cys	
850 855 860 865	
ttt att ctc tta gct aga act att gat caa aaa gtg aaa aat gct ttt	2926
Phe Ile Leu Leu Ala Arg Thr Ile Asp Gln Lys Val Lys Asn Ala Phe	
870 875 880	
cca att gct att atc tgg att act ttg aca ttg ttt tat ctt aat tta	2974
Pro Ile Ala Ile Ile Trp Ile Thr Leu Thr Leu Phe Tyr Leu Asn Leu	
885 890 895	
ggg cat att agt tgg cga cta tct ttc tgg ttt att tta cta ttg tta	3022
Gly His Ile Ser Trp Arg Leu Ser Phe Trp Phe Ile Leu Leu Leu Leu	
900 905 910	
ggc tta tta gtc att aag cca act ctc tat aaa aaa caa ttt att tat	3070
Gly Leu Leu Val Ile Lys Pro Thr Leu Tyr Lys Lys Gln Phe Ile Tyr	
915 920 925	
agc tgg gaa gag cgt att aag gat gga atc att atc gtt agt tta atg	3118
Ser Trp Glu Glu Arg Ile Lys Asp Gly Ile Ile Ile Val Ser Leu Met	
930 935 940 945	
gga gtt cta ttt tat att gca gga cta cta ttc cct atc agg gct cat	3166
Gly Val Leu Phe Tyr Ile Ala Gly Leu Leu Phe Pro Ile Arg Ala His	
950 955 960	
att aca ggt ggt agt att gaa cgc ctg cat tat atc ata gca tgg gag	3214
Ile Thr Gly Gly Ser Ile Glu Arg Leu His Tyr Ile Ile Ala Trp Glu	
965 970 975	
ccg ata gca ttg gct acg ttg att ctt act ctc gtt tat tta tgt ttg	3262
Pro Ile Ala Leu Ala Thr Leu Ile Leu Thr Leu Val Tyr Leu Cys Leu	
980 985 990	
ggt aag att tta caa gga aaa tct tgt cag att ggt gat gtg ttc aat	3310
Val Lys Ile Leu Gln Gly Lys Ser Cys Gln Ile Gly Asp Val Phe Asn	
995 1000 1005	
gtg gat cgt tat aaa aaa cta ctt caa gct tac ggt ggt tct tcg gat	3358
Val Asp Arg Tyr Lys Lys Leu Leu Gln Ala Tyr Gly Gly Ser Ser Asp	
1010 1015 1020 1025	
agc ggt tta gcc ttt tta aat gat aaa agg ctc tac tgg tac caa aaa	3406
Ser Gly Leu Ala Phe Leu Asn Asp Lys Arg Leu Tyr Trp Tyr Gln Lys	
1030 1035 1040	
aat gga gaa gat tgc gtt gcg ttc caa ttt gta att gtc aat aat aaa	3454
Asn Gly Glu Asp Cys Val Ala Phe Gln Phe Val Ile Val Asn Asn Lys	
1045 1050 1055	
tgt ctt att atg ggg gaa cca gcc ggt gat gac act tat att cgt gaa	3502
Cys Leu Ile Met Gly Glu Pro Ala Gly Asp Asp Thr Tyr Ile Arg Glu	
1060 1065 1070	

7 / 63

gct att gaa tcg ttt att gat gat gct gat aag cta gac tat gac ctt 3550
Ala Ile Glu Ser Phe Ile Asp Asp Ala Asp Lys Leu Asp Tyr Asp Leu
1075 1080 1085

gtt ttt tac agt att gga cag aag ttg aca cta ctt tta cat gag tat 3598
Val Phe Tyr Ser Ile Gly Gln Lys Leu Thr Leu Leu Leu His Glu Tyr
1090 1095 1100 1105

ggt ttt gac ttt atg aaa gtt ggt gag gat gct tta gtt aat tta gaa 3646
Gly Phe Asp Phe Met Lys Val Gly Glu Asp Ala Leu Val Asn Leu Glu
1110 1115 1120

acg ttt act ctt aaa ggg aat aag tac aaa cct ttc aga aat gcc cta 3694
Thr Phe Thr Leu Lys Gly Asn Lys Tyr Lys Pro Phe Arg Asn Ala Leu
1125 1130 1135

aat aga gtt gaa aag gat ggt ttc tat ttc gaa gtt gta caa tcg cca 3742
Asn Arg Val Glu Lys Asp Gly Phe Tyr Phe Glu Val Val Gln Ser Pro
1140 1145 1150

cat agt caa gag cta cta aat agt ttg gaa gag att tct aat act tgg 3790
His Ser Gln Glu Leu Leu Asn Ser Leu Glu Glu Ile Ser Asn Thr Trp
1155 1160 1165

tta gaa gga cgt cct gaa aaa ggt ttc tca cta gga tat ttt aat aaa 3838
Leu Glu Gly Arg Pro Glu Lys Gly Phe Ser Leu Gly Tyr Phe Asn Lys
1170 1175 1180 1185

gat tat ttc caa caa gcc cca ata gct ttg gta aaa aat gct gaa cac 3886
Asp Tyr Phe Gln Ala Pro Ile Ala Leu Val Lys Asn Ala Glu His
1190 1195 1200

gaa gtt gtt gct ttt gct aat att atg cca aac tat gaa aag agt att 3934
Glu Val Val Ala Phe Ala Asn Ile Met Pro Asn Tyr Glu Lys Ser Ile
1205 1210 1215

atc tct att gat tta atg cgt cac gat aaa cag aaa att ccg aat ggc 3982
Ile Ser Ile Asp Leu Met Arg His Asp Lys Gln Lys Ile Pro Asn Gly
1220 1225 1230

gtt atg gat ttc ctc ttt tta tca tta ttc tct tat tat caa gag aag 4030
Val Met Asp Phe Leu Phe Leu Ser Leu Phe Ser Tyr Tyr Gln Glu Lys
1235 1240 1245

gga tac cac tat ttt gat ttg ggg atg gca cct tta tca gga gtt ggt 4078
Gly Tyr His Tyr Phe Asp Leu Gly Met Ala Pro Leu Ser Gly Val Gly
1250 1255 1260 1265

cgc gtt gaa aca agt ttt gct aaa gag aga atg gcg tat ctt gtc tat 4126
Arg Val Glu Thr Ser Phe Ala Lys Glu Arg Met Ala Tyr Leu Val Tyr
1270 1275 1280

cat ttc ggt agt cat ttc tac tca ttt aat ggt tta cac aag tat aag 4174
His Phe Gly Ser His Phe Tyr Ser Phe Asn Gly Leu His Lys Tyr Lys
1285 1290 1295

aag aag ttt aca cca ttg tgg tcg gaa cgt tat att tct tgt tct cgt 4222
 Lys Lys Phe Thr Pro Leu Trp Ser Glu Arg Tyr Ile Ser Cys Ser Arg
 1300 1305 1310

tcg tcc tgg tta att tgt gct att tgt gcc cta tta atg gaa gat agt 4270
 Ser Ser Trp Leu Ile Cys Ala Ile Cys Ala Leu Leu Met Glu Asp Ser
 1315 1320 1325

aaa att aag att gtt aaa taagctttat ttggcaatta aaaagagcat 4318
 Lys Ile Lys Ile Val Lys
 1330 1335

gtcatgcgac atgctctttt taaatcattt aataccattg attgcttgaa tctactttat 4378
 aatatgatgt gctttttaaatt attgttttagc tactgtagct gctgatttat gctttacagc 4438
 tacttggtag ttcattttctt gcattttcttt ttcagtgata tgaccagcaa gttttattgag 4498
 agctttttttt acttga 4514

<210> 2
 <211> 154
 <212> PRT
 <213> Streptococcus

<400> 2
 Ser Gly Lys Glu Pro Ala Asn Arg Phe Ser Trp Ala Lys Asn Lys Leu
 1 5 10 15
 Leu Ile Asn Gly Phe Ile Ala Thr Leu Ala Ala Thr Ile Leu Phe Phe
 20 25 30
 Ala Val Gln Phe Ile Gly Leu Lys Pro Asp Tyr Pro Gly Lys Thr Tyr
 35 40 45
 Phe Ile Ile Leu Leu Thr Ala Trp Thr Leu Met Ala Leu Val Thr Ala
 50 55 60
 Leu Val Gly Trp Asp Asn Arg Tyr Gly Ser Phe Leu Ser Leu Leu Ile
 65 70 75 80
 Leu Leu Phe Gln Leu Gly Ser Ser Ala Gly Thr Tyr Pro Ile Glu Leu
 85 90 95
 Ser Pro Lys Phe Phe Gln Thr Ile Gln Pro Phe Leu Pro Met Thr Tyr
 100 105 110
 Ser Val Ser Gly Leu Arg Glu Thr Ile Ser Leu Thr Gly Asp Val Asn
 115 120 125
 His Gln Trp Arg Met Leu Val Ile Phe Leu Val Ser Ser Met Ile Leu
 130 135 140
 Ala Leu Leu Ile Tyr Arg Lys Gln Glu Asp
 145 150

<210> 3
 <211> 118
 <212> PRT
 <213> Streptococcus

<400> 3
 Met Ser Thr Leu Thr Ile Ile Ile Ala Thr Leu Thr Ala Leu Glu His
 1 5 10 15
 Phe Tyr Ile Met Tyr Leu Glu Thr Leu Ala Thr Gln Ser Asn Met Thr
 20 25 30

Gly Lys Ile Phe Ser Met Ser Lys Glu Glu Leu Ser Tyr Leu Pro Val
 35 40 45
 Ile Lys Leu Phe Lys Asn Gln Gly Val Tyr Asn Gly Leu Ile Gly Leu
 50 55 60
 Phe Leu Leu Tyr Gly Leu Tyr Ile Ser Gln Asn Gln Glu Ile Val Ala
 65 70 75 80
 Val Phe Leu Ile Asn Val Leu Leu Val Ala Ile Tyr Gly Ala Leu Thr
 85 90 95
 Val Asp Lys Lys Ile Leu Leu Lys Gln Gly Gly Leu Pro Ile Leu Ala
 100 105 110
 Leu Leu Thr Phe Leu Phe
 115

<210> 4
 <211> 247
 <212> PRT
 <213> Streptococcus

<400> 4
 Met Thr Glu Asn Trp Leu His Thr Lys Asp Gly Ser Asp Ile Tyr Tyr
 1 5 10 15
 Arg Val Val Gly Gln Gly Gln Pro Ile Val Phe Leu His Gly Asn Ser
 20 25 30
 Leu Ser Ser Arg Tyr Phe Asp Lys Gln Ile Ala Tyr Phe Ser Lys Tyr
 35 40 45
 Tyr Gln Val Ile Val Met Asp Ser Arg Gly His Gly Lys Ser His Ala
 50 55 60
 Lys Leu Asn Thr Ile Ser Phe Arg Gln Ile Ala Val Asp Leu Lys Asp
 65 70 75 80
 Ile Leu Val His Leu Glu Ile Asp Lys Val Ile Leu Val Gly His Ser
 85 90 95
 Asp Gly Ala Asn Leu Ala Leu Val Phe Gln Thr Met Phe Pro Gly Met
 100 105 110
 Val Arg Gly Leu Leu Leu Asn Ser Gly Asn Leu Thr Ile His Gly Gln
 115 120 125
 Arg Trp Trp Asp Ile Leu Leu Val Arg Ile Ala Tyr Lys Phe Leu His
 130 135 140
 Tyr Leu Gly Lys Leu Phe Pro Tyr Met Arg Gln Lys Ala Gln Val Ile
 145 150 155 160
 Ser Leu Met Leu Glu Asp Leu Lys Ile Ser Pro Ala Asp Leu Gln His
 165 170 175
 Val Ser Thr Pro Val Met Val Leu Val Gly Asn Lys Asp Ile Ile Lys
 180 185 190
 Leu Asn His Ser Lys Lys Leu Ala Ser Tyr Phe Pro Arg Gly Glu Phe
 195 200 205
 Tyr Ser Leu Val Gly Phe Gly His His Ile Ile Lys Gln Asp Ser His
 210 215 220
 Val Phe Asn Ile Ile Ala Lys Lys Phe Ile Asn Asp Thr Leu Lys Gly
 225 230 235 240
 Glu Ile Val Glu Lys Ala Asn
 245

<210> 5
 <211> 816
 <212> PRT
 <213> Streptococcus

<400> 5
 Met Ile His Leu Lys Arg Thr Ile Ser Val Glu Gln Leu Lys Ser Val
 1 5 10 15
 Phe Gly Gln Leu Ser Pro Met Asn Leu Phe Leu Ile Ile Leu Val Gly
 20 25 30
 Val Ile Ala Val Leu Pro Thr Thr Gly Tyr Asp Phe Val Leu Asn Gly
 35 40 45
 Leu Leu Arg Thr Asp Lys Ser Lys Arg Tyr Ile Leu Gln Thr Ser Trp
 50 55 60
 Cys Ile Asn Thr Phe Asn Asn Leu Ser Gly Phe Gly Gly Leu Ile Asp
 65 70 75 80
 Ile Gly Leu Arg Met Ala Phe Tyr Gly Lys Lys Gly Gln Glu Lys Ser
 85 90 95
 Asp Leu Arg Glu Val Thr Arg Phe Leu Pro Tyr Leu Ile Ser Gly Leu
 100 105 110
 Ser Phe Ile Ser Val Ile Ala Leu Ile Met Ser His Ile Phe His Ala
 115 120 125
 Lys Ala Ser Val Asp Tyr Tyr Tyr Leu Val Leu Ile Gly Ala Ser Met
 130 135 140
 Tyr Phe Pro Val Ile Tyr Trp Ile Ser Gly His Lys Gly Ser His Tyr
 145 150 155 160
 Phe Gly Asp Met Pro Ser Ser Thr Arg Ile Lys Leu Gly Val Val Ser
 165 170 175
 Phe Phe Glu Trp Gly Cys Ala Ala Ala Phe Ile Ile Ile Gly Tyr
 180 185 190
 Leu Met Gly Ile His Leu Pro Val Tyr Lys Ile Leu Pro Leu Phe Cys
 195 200 205
 Ile Gly Cys Ala Val Gly Ile Val Ser Leu Ile Pro Gly Gly Leu Gly
 210 215 220
 Ser Phe Glu Leu Val Leu Phe Thr Gly Phe Ala Ala Glu Gly Leu Pro
 225 230 235 240
 Lys Glu Thr Val Val Ala Trp Leu Leu Leu Tyr Arg Leu Ala Tyr Tyr
 245 250 255
 Ile Ile Pro Phe Phe Ala Gly Ile Tyr Phe Phe Ile His Tyr Leu Gly
 260 265 270
 Ser Gln Ile Asn Gln Arg Tyr Glu Asn Val Pro Lys Glu Leu Val Ser
 275 280 285
 Thr Val Leu Gln Thr Met Val Ser His Leu Met Arg Ile Leu Gly Ala
 290 295 300
 Phe Leu Ile Phe Ser Thr Ala Phe Phe Glu Asn Ile Thr Tyr Ile Met
 305 310 315 320
 Trp Leu Gln Lys Leu Gly Leu Asp Pro Leu Gln Glu Gln Met Leu Trp
 325 330 335
 Gln Phe Pro Gly Leu Leu Leu Gly Val Cys Phe Ile Leu Leu Ala Arg
 340 345 350
 Thr Ile Asp Gln Lys Val Lys Asn Ala Phe Pro Ile Ala Ile Ile Trp
 355 360 365
 Ile Thr Leu Thr Leu Phe Tyr Leu Asn Leu Gly His Ile Ser Trp Arg
 370 375 380
 Leu Ser Phe Trp Phe Ile Leu Leu Leu Leu Gly Leu Leu Val Ile Lys
 385 390 395 400
 Pro Thr Leu Tyr Lys Lys Gln Phe Ile Tyr Ser Trp Glu Glu Arg Ile
 405 410 415
 Lys Asp Gly Ile Ile Ile Val Ser Leu Met Gly Val Leu Phe Tyr Ile
 420 425 430

Ala Gly Leu Leu Phe Pro Ile Arg Ala His Ile Thr Gly Gly Ser Ile
 435 440 445
 Glu Arg Leu His Tyr Ile Ile Ala Trp Glu Pro Ile Ala Leu Ala Thr
 450 455 460
 Leu Ile Leu Thr Leu Val Tyr Leu Cys Leu Val Lys Ile Leu Gln Gly
 465 470 475 480
 Lys Ser Cys Gln Ile Gly Asp Val Phe Asn Val Asp Arg Tyr Lys Lys
 485 490 495
 Leu Leu Gln Ala Tyr Gly Gly Ser Ser Asp Ser Gly Leu Ala Phe Leu
 500 505 510
 Asn Asp Lys Arg Leu Tyr Trp Tyr Gln Lys Asn Gly Glu Asp Cys Val
 515 520 525
 Ala Phe Gln Phe Val Ile Val Asn Asn Lys Cys Leu Ile Met Gly Glu
 530 535 540
 Pro Ala Gly Asp Asp Thr Tyr Ile Arg Glu Ala Ile Glu Ser Phe Ile
 545 550 555 560
 Asp Asp Ala Asp Lys Leu Asp Tyr Asp Leu Val Phe Tyr Ser Ile Gly
 565 570 575
 Gln Lys Leu Thr Leu Leu Leu His Glu Tyr Gly Phe Asp Phe Met Lys
 580 585 590
 Val Gly Glu Asp Ala Leu Val Asn Leu Glu Thr Phe Thr Leu Lys Gly
 595 600 605
 Asn Lys Tyr Lys Pro Phe Arg Asn Ala Leu Asn Arg Val Glu Lys Asp
 610 615 620
 Gly Phe Tyr Phe Glu Val Val Gln Ser Pro His Ser Gln Glu Leu Leu
 625 630 635 640
 Asn Ser Leu Glu Glu Ile Ser Asn Thr Trp Leu Glu Gly Arg Pro Glu
 645 650 655
 Lys Gly Phe Ser Leu Gly Tyr Phe Asn Lys Asp Tyr Phe Gln Gln Ala
 660 665 670
 Pro Ile Ala Leu Val Lys Asn Ala Glu His Glu Val Val Ala Phe Ala
 675 680 685
 Asn Ile Met Pro Asn Tyr Glu Lys Ser Ile Ile Ser Ile Asp Leu Met
 690 695 700
 Arg His Asp Lys Gln Lys Ile Pro Asn Gly Val Met Asp Phe Leu Phe
 705 710 715 720
 Leu Ser Leu Phe Ser Tyr Tyr Gln Glu Lys Gly Tyr His Tyr Phe Asp
 725 730 735
 Leu Gly Met Ala Pro Leu Ser Gly Val Gly Arg Val Glu Thr Ser Phe
 740 745 750
 Ala Lys Glu Arg Met Ala Tyr Leu Val Tyr His Phe Gly Ser His Phe
 755 760 765
 Tyr Ser Phe Asn Gly Leu His Lys Tyr Lys Lys Lys Phe Thr Pro Leu
 770 775 780
 Trp Ser Glu Arg Tyr Ile Ser Cys Ser Arg Ser Ser Trp Leu Ile Cys
 785 790 795 800
 Ala Ile Cys Ala Leu Leu Met Glu Asp Ser Lys Ile Lys Ile Val Lys
 805 810 815

<210> 6
 <211> 518
 <212> PRT
 <213> Streptococcus

<400> 6

12 / 63

Met Arg Ile Leu Gly Ala Phe Leu Ile Phe Ser Thr Ala Phe Phe Glu
 1 5 10 15
 Asn Ile Thr Tyr Ile Met Trp Leu Gln Lys Leu Gly Leu Asp Pro Leu
 20 25 30
 Gln Glu Gln Met Leu Trp Gln Phe Pro Gly Leu Leu Leu Gly Val Cys
 35 40 45
 Phe Ile Leu Leu Ala Arg Thr Ile Asp Gln Lys Val Lys Asn Ala Phe
 50 55 60
 Pro Ile Ala Ile Ile Trp Ile Thr Leu Thr Leu Phe Tyr Leu Asn Leu
 65 70 75 80
 Gly His Ile Ser Trp Arg Leu Ser Phe Trp Phe Ile Leu Leu Leu Leu
 85 90 95
 Gly Leu Leu Val Ile Lys Pro Thr Leu Tyr Lys Lys Gln Phe Ile Tyr
 100 105 110
 Ser Trp Glu Glu Arg Ile Lys Asp Gly Ile Ile Ile Val Ser Leu Met
 115 120 125
 Gly Val Leu Phe Tyr Ile Ala Gly Leu Leu Phe Pro Ile Arg Ala His
 130 135 140
 Ile Thr Gly Gly Ser Ile Glu Arg Leu His Tyr Ile Ile Ala Trp Glu
 145 150 155 160
 Pro Ile Ala Leu Ala Thr Leu Ile Leu Thr Leu Val Tyr Leu Cys Leu
 165 170 175
 Val Lys Ile Leu Gln Gly Lys Ser Cys Gln Ile Gly Asp Val Phe Asn
 180 185 190
 Val Asp Arg Tyr Lys Lys Leu Leu Gln Ala Tyr Gly Gly Ser Ser Asp
 195 200 205
 Ser Gly Leu Ala Phe Leu Asn Asp Lys Arg Leu Tyr Trp Tyr Gln Lys
 210 215 220
 Asn Gly Glu Asp Cys Val Ala Phe Gln Phe Val Ile Val Asn Asn Lys
 225 230 235 240
 Cys Leu Ile Met Gly Glu Pro Ala Gly Asp Asp Thr Tyr Ile Arg Glu
 245 250 255
 Ala Ile Glu Ser Phe Ile Asp Asp Ala Asp Lys Leu Asp Tyr Asp Leu
 260 265 270
 Val Phe Tyr Ser Ile Gly Gln Lys Leu Thr Leu Leu Leu His Glu Tyr
 275 280 285
 Gly Phe Asp Phe Met Lys Val Gly Glu Asp Ala Leu Val Asn Leu Glu
 290 295 300
 Thr Phe Thr Leu Lys Gly Asn Lys Tyr Lys Pro Phe Arg Asn Ala Leu
 305 310 315 320
 Asn Arg Val Glu Lys Asp Gly Phe Tyr Phe Glu Val Val Gln Ser Pro
 325 330 335
 His Ser Gln Glu Leu Leu Asn Ser Leu Glu Glu Ile Ser Asn Thr Trp
 340 345 350
 Leu Glu Gly Arg Pro Glu Lys Gly Phe Ser Leu Gly Tyr Phe Asn Lys
 355 360 365
 Asp Tyr Phe Gln Gln Ala Pro Ile Ala Leu Val Lys Asn Ala Glu His
 370 375 380
 Glu Val Val Ala Phe Ala Asn Ile Met Pro Asn Tyr Glu Lys Ser Ile
 385 390 395 400
 Ile Ser Ile Asp Leu Met Arg His Asp Lys Gln Lys Ile Pro Asn Gly
 405 410 415
 Val Met Asp Phe Leu Phe Leu Ser Leu Phe Ser Tyr Tyr Gln Glu Lys
 420 425 430
 Gly Tyr His Tyr Phe Asp Leu Gly Met Ala Pro Leu Ser Gly Val Gly
 435 440 445

Arg Val Glu Thr Ser Phe Ala Lys Glu Arg Met Ala Tyr Leu Val Tyr
 450 455 460
 His Phe Gly Ser His Phe Tyr Ser Phe Asn Gly Leu His Lys Tyr Lys
 465 470 475 480
 Lys Lys Phe Thr Pro Leu Trp Ser Glu Arg Tyr Ile Ser Cys Ser Arg
 485 490 495
 Ser Ser Trp Leu Ile Cys Ala Ile Cys Ala Leu Leu Met Glu Asp Ser
 500 505 510
 Lys Ile Lys Ile Val Lys
 515

<210> 7
 <211> 5126
 <212> DNA
 <213> Streptococcus

<220>
 <221> CDS
 <222> (1)...(687)
 <221> CDS
 <222> (701)...(2557)

<221> CDS
 <222> (2566)...(3036)
 <221> CDS
 <222> (3106)...(4842)
 <221> CDS
 <222> (4850)...(5125)

<400> 7
 aat ttt gat atc gaa aca aca act ttt gag gca atg aaa aag cac gcg 48
 Asn Phe Asp Ile Glu Thr Thr Thr Phe Glu Ala Met Lys Lys His Ala
 1 5 10 15
 tca tta ttg gag aaa ata tct gtt gag cgt tct ttt att gaa ttt gat 96
 Ser Leu Leu Glu Lys Ile Ser Val Glu Arg Ser Phe Ile Glu Phe Asp
 20 25 30
 aaa ctt cta tta gca cct tat tgg cgt aaa gga atg ctg gca cta ata 144
 Lys Leu Leu Leu Ala Pro Tyr Trp Arg Lys Gly Met Leu Ala Leu Ile
 35 40 45
 gat agt cat gct ttt aat tat cta cca tgc tta aaa aat agg gaa tta 192
 Asp Ser His Ala Phe Asn Tyr Leu Pro Cys Leu Lys Asn Arg Glu Leu
 50 55 60
 caa tta agc gcc ttt ttg tcc cag tta gat aaa gat ttt tta ttt gag 240
 Gln Leu Ser Ala Phe Leu Ser Gln Leu Asp Lys Asp Phe Leu Phe Glu
 65 70 75 80
 aca tca gaa caa gct tgg gca tca ctc atc ttg agt atg gaa gtt gaa 288
 Thr Ser Glu Gln Ala Trp Ala Ser Leu Ile Leu Ser Met Glu Val Glu
 85 90 95

cac	aca	aag	act	ttt	tta	aaa	aaa	tgg	aag	aca	tca	act	cac	ttt	caa	336
His	Thr	Lys	Thr	Phe	Leu	Lys	Lys	Trp	Lys	Thr	Ser	Thr	His	Phe	Gln	
			100					105					110			
aaa	gat	ggt	gag	cat	ata	gtg	gat	ggt	tat	cgt	att	cgt	gaa	caa	atg	384
Lys	Asp	Val	Glu	His	Ile	Val	Asp	Val	Tyr	Arg	Ile	Arg	Glu	Gln	Met	
		115					120					125				
gga	ttg	gct	aaa	gaa	cat	ctt	tat	cgt	tat	gga	aaa	act	ata	ata	aaa	432
Gly	Leu	Ala	Lys	Glu	His	Leu	Tyr	Arg	Tyr	Gly	Lys	Thr	Ile	Ile	Lys	
	130					135					140					
caa	gcg	gaa	ggt	att	cgc	aaa	gca	aga	ggc	ttg	atg	ggt	gat	ttc	gaa	480
Gln	Ala	Glu	Gly	Ile	Arg	Lys	Ala	Arg	Gly	Leu	Met	Val	Asp	Phe	Glu	
145					150					155					160	
aaa	ata	gaa	caa	cta	gat	agt	gag	tta	gca	atc	cat	gat	agg	cat	gag	528
Lys	Ile	Glu	Gln	Leu	Asp	Ser	Glu	Leu	Ala	Ile	His	Asp	Arg	His	Glu	
				165					170					175		
ata	ggt	gtc	aat	ggt	ggc	acc	tta	atc	aag	aaa	tta	gga	ata	aaa	cct	576
Ile	Val	Val	Asn	Gly	Gly	Thr	Leu	Ile	Lys	Lys	Leu	Gly	Ile	Lys	Pro	
			180					185					190			
ggt	cca	cag	atg	gga	gat	att	atc	tct	caa	att	gaa	tta	gcc	att	ggt	624
Gly	Pro	Gln	Met	Gly	Asp	Ile	Ile	Ser	Gln	Ile	Glu	Leu	Ala	Ile	Val	
		195				200						205				
tta	gga	caa	ctg	att	aat	gaa	gaa	gag	gct	att	tta	cat	ttt	ggt	aag	672
Leu	Gly	Gln	Leu	Ile	Asn	Glu	Glu	Glu	Ala	Ile	Leu	His	Phe	Val	Lys	
	210				215						220					
cag	tac	ttg	atg	gat	tagagaggat	tat	atg	agc	gat	ttt	tta	gta	gat			721
Gln	Tyr	Leu	Met	Asp			Met	Ser	Asp	Phe	Leu	Val	Asp			
225							230					235				
gga	ttg	act	aag	tcg	ggt	ggt	gat	aag	acg	gtc	ttt	agt	aat	ggt	tca	769
Gly	Leu	Thr	Lys	Ser	Val	Gly	Asp	Lys	Thr	Val	Phe	Ser	Asn	Val	Ser	
			240				245						250			
ttt	atc	atc	cat	agt	tta	gac	cgt	att	ggg	att	att	ggt	gtc	aat	gga	817
Phe	Ile	Ile	His	Ser	Leu	Asp	Arg	Ile	Gly	Ile	Ile	Gly	Val	Asn	Gly	
		255				260						265				
act	gga	aag	aca	aca	cta	tta	gat	ggt	att	tcg	ggt	gaa	tta	ggt	ttt	865
Thr	Gly	Lys	Thr	Thr	Leu	Leu	Asp	Val	Ile	Ser	Gly	Glu	Leu	Gly	Phe	
	270					275					280					
gat	ggt	gat	cgt	tcc	cct	ttt	tca	tca	gct	aat	gat	tat	aag	att	gct	913
Asp	Gly	Asp	Arg	Ser	Pro	Phe	Ser	Ser	Ala	Asn	Asp	Tyr	Lys	Ile	Ala	
285					290					295					300	
tat	tta	aaa	caa	gaa	cca	gac	ttt	gat	gat	tct	cag	aca	att	ttg	gac	961
Tyr	Leu	Lys	Gln	Glu	Pro	Asp	Phe	Asp	Asp	Ser	Gln	Thr	Ile	Leu	Asp	
			305					310						315		

acc gta ctt tct tct gac tta aga gag atg gct tta att aaa gaa tat 1009
 Thr Val Leu Ser Ser Asp Leu Arg Glu Met Ala Leu Ile Lys Glu Tyr
 320 325 330

gaa tta ttg ctt aat cac tac gaa gaa agt aag caa tca cgt cta gag 1057
 Glu Leu Leu Leu Asn His Tyr Glu Glu Ser Lys Gln Ser Arg Leu Glu
 335 340 345

aaa gta atg gca gaa atg gat tct tta gat gct tgg tct att gag agc 1105
 Lys Val Met Ala Glu Met Asp Ser Leu Asp Ala Trp Ser Ile Glu Ser
 350 355 360

gaa gtc aaa aca gta tta tcc aaa tta ggt att act gat ttg cag ttg 1153
 Glu Val Lys Thr Val Leu Ser Lys Leu Gly Ile Thr Asp Leu Gln Leu
 365 370 375 380

tcg gtt ggt gaa tta tca gga gga tta cga aga cgt gtt caa tta gcg 1201
 Ser Val Gly Glu Leu Ser Gly Gly Leu Arg Arg Arg Val Gln Leu Ala
 385 390 395

caa gta tta tta aat gat gca gat tta ttg ctc tta gac gaa cct act 1249
 Gln Val Leu Leu Asn Asp Ala Asp Leu Leu Leu Leu Asp Glu Pro Thr
 400 405 410

aac cac tta gat att gac act att gca tgg tta acg aat ttt ttg aaa 1297
 Asn His Leu Asp Ile Asp Thr Ile Ala Trp Leu Thr Asn Phe Leu Lys
 415 420 425

aat agt aaa aag aca gtg ctt ttt ata act cat gat cgt tat ttt cta 1345
 Asn Ser Lys Lys Thr Val Leu Phe Ile Thr His Asp Arg Tyr Phe Leu
 430 435 440

gac aat gtt gca aca cgt att ttt gaa tta gat aag gca cag att aca 1393
 Asp Asn Val Ala Thr Arg Ile Phe Glu Leu Asp Lys Ala Gln Ile Thr
 445 450 455 460

gaa tat caa ggc aat tat cag gat tat gtc cga ctt cgt gca gaa caa 1441
 Glu Tyr Gln Gly Asn Tyr Gln Asp Tyr Val Arg Leu Arg Ala Glu Gln
 465 470 475

gac gag cgt gat gct gct agt tta cat aaa aag aaa cag ctt tat aaa 1489
 Asp Glu Arg Asp Ala Ala Ser Leu His Lys Lys Lys Gln Leu Tyr Lys
 480 485 490

cag gaa cta gct tgg atg cgt act cag cca caa gct cgt gca acg aaa 1537
 Gln Glu Leu Ala Trp Met Arg Thr Gln Pro Gln Ala Arg Ala Thr Lys
 495 500 505

caa cag gct cgt att aat cgt ttt caa aat cta aaa aac gat tta cac 1585
 Gln Gln Ala Arg Ile Asn Arg Phe Gln Asn Leu Lys Asn Asp Leu His
 510 515 520

caa aca agc gat aca agc gat ttg gaa atg aca ttt gaa aca agt cga 1633
 Gln Thr Ser Asp Thr Ser Asp Leu Glu Met Thr Phe Glu Thr Ser Arg
 525 530 535 540

16 / 63

att ggg aaa aag gtt att aat ttt gaa aat gtc tct ttt tct tac cca	1681
Ile Gly Lys Lys Val Ile Asn Phe Glu Asn Val Ser Phe Ser Tyr Pro	
545 550 555	
gat aaa tct atc ttg aaa gac ttt aat ttg tta att caa aat aaa gac	1729
Asp Lys Ser Ile Leu Lys Asp Phe Asn Leu Leu Ile Gln Asn Lys Asp	
560 565 570	
cgt att ggc atc gtt gga gat aat ggt gtt gga aag tca acc tta ctt	1777
Arg Ile Gly Ile Val Gly Asp Asn Gly Val Gly Lys Ser Thr Leu Leu	
575 580 585	
aat tta att gtt caa gat tta cag ccg gat tcg ggt aat gtc tct att	1825
Asn Leu Ile Val Gln Asp Leu Gln Pro Asp Ser Gly Asn Val Ser Ile	
590 595 600	
ggt gaa acg ata cgt gta ggt tac ttt tca caa caa ctt cat aat atg	1873
Gly Glu Thr Ile Arg Val Gly Tyr Phe Ser Gln Gln Leu His Asn Met	
605 610 615 620	
gat ggc tca aaa cgt gtt att aat tat ttg caa gag gtt gca gat gag	1921
Asp Gly Ser Lys Arg Val Ile Asn Tyr Leu Gln Glu Val Ala Asp Glu	
625 630 635	
gtt aaa act agt gtc ggt aca aca agt gtg aca gaa cta ttg gaa caa	1969
Val Lys Thr Ser Val Gly Thr Thr Ser Val Thr Glu Leu Leu Glu Gln	
640 645 650	
ttt ctc ttt cca cgt tcg aca cat gga aca caa att gca aaa tta tca	2017
Phe Leu Phe Pro Arg Ser Thr His Gly Thr Gln Ile Ala Lys Leu Ser	
655 660 665	
ggt ggt gag aaa aaa aga ctt tac ctt tta aaa atc ctg att gaa aag	2065
Gly Gly Glu Lys Lys Arg Leu Tyr Leu Leu Lys Ile Leu Ile Glu Lys	
670 675 680	
cct aat gtg tta cta ctt gat gag ccg aca aat gac tta gat att gct	2113
Pro Asn Val Leu Leu Leu Asp Glu Pro Thr Asn Asp Leu Asp Ile Ala	
685 690 695 700	
aca tta act gtt ctt gaa aat ttt tta caa ggc ttt ggt ggt cct gtg	2161
Thr Leu Thr Val Leu Glu Asn Phe Leu Gln Gly Phe Gly Gly Pro Val	
705 710 715	
att aca gtt agt cac gat cgt tac ttt tta gat aaa gtg gct aat aaa	2209
Ile Thr Val Ser His Asp Arg Tyr Phe Leu Asp Lys Val Ala Asn Lys	
720 725 730	
att att gcg ttt gaa gat aac gat atc cgt gaa ttt ttt ggt aat tat	2257
Ile Ile Ala Phe Glu Asp Asn Asp Ile Arg Glu Phe Phe Gly Asn Tyr	
735 740 745	
act gat tat tta gat gaa aaa gca ttt aat gag caa aat aat gaa gtt	2305
Thr Asp Tyr Leu Asp Glu Lys Ala Phe Asn Glu Gln Asn Asn Glu Val	
750 755 760	

atc agt aaa aaa gag agt acc aag aca agt cgt gaa aag caa agt cgt Ile Ser Lys Lys Glu Ser Thr Lys Thr Ser Arg Glu Lys Gln Ser Arg 765 770 775 780	2353
aaa aga atg tct tac ttt gaa aaa caa gaa tgg gcg aca att gaa gac Lys Arg Met Ser Tyr Phe Glu Lys Gln Glu Trp Ala Thr Ile Glu Asp 785 790 795	2401
gat att atg ata ttg gaa aat act atc act cgt ata gaa aat gat atg Asp Ile Met Ile Leu Glu Asn Thr Ile Thr Arg Ile Glu Asn Asp Met 800 805 810	2449
caa aca tgt ggt agt gat ttt aca agg tta tct gat tta caa aag gaa Gln Thr Cys Gly Ser Asp Phe Thr Arg Leu Ser Asp Leu Gln Lys Glu 815 820 825	2497
tta gat gca aaa aat gaa gca ctt cta gaa aag tat gac cgt tat gag Leu Asp Ala Lys Asn Glu Ala Leu Leu Glu Lys Tyr Asp Arg Tyr Glu 830 835 840	2545
tac ctt agt gag ttagacac atg att atc cgt ccg att att aaa aat gat Tyr Leu Ser Glu LeuAspThrMet Ile Ile Arg Pro Ile Ile Lys Asn Asp 845 850 855 860	2595
gac caa gca gtt gca caa tta att cga caa agt tta cgc gcc tat gat Asp Gln Ala Val Ala Gln Leu Ile Arg Gln Ser Leu Arg Ala Tyr Asp 865 870 875	2643
tta gat aaa cct gat aca gca tat tca gac cct cac tta gat cat ttg Leu Asp Lys Pro Asp Thr Ala Tyr Ser Asp Pro His Leu Asp His Leu 880 885 890	2691
acc tca tac tac gaa aaa ata gag aag tca gga ttc ttt gtc att gag Thr Ser Tyr Tyr Glu Lys Ile Glu Lys Ser Gly Phe Phe Val Ile Glu 895 900 905	2739
gag aga gat gag att att ggc tgt ggc ggc ttt ggt ccg ctg aaa aat Glu Arg Asp Glu Ile Ile Gly Cys Gly Gly Phe Gly Pro Leu Lys Asn 910 915 920 925	2787
cta att gca gag atg cag aag gtg tac att gca gaa cgt ttc cgt ggt Leu Ile Ala Glu Met Gln Lys Val Tyr Ile Ala Glu Arg Phe Arg Gly 930 935 940	2835
aag ggg ctt gct act gat tta gtg aaa atg att gaa gta gaa gct cga Lys Gly Leu Ala Thr Asp Leu Val Lys Met Ile Glu Val Glu Ala Arg 945 950 955	2883
aaa att ggg tat aga caa ctt tat tta gag aca gcc agt act ttg agt Lys Ile Gly Tyr Arg Gln Leu Tyr Leu Glu Thr Ala Ser Thr Leu Ser 960 965 970	2931
agg gca act gcg gtt tat aag cat atg gga tat tgt gcc tta tcg caa Arg Ala Thr Ala Val Tyr Lys His Met Gly Tyr Cys Ala Leu Ser Gln 975 980 985	2979

cca ata gca aat gat caa ggt cat aca gct atg gat att tgg atg att 3027
Pro Ile Ala Asn Asp Gln Gly His Thr Ala Met Asp Ile Trp Met Ile
990 995 1000 1005

aaa gat tta taagttgaaa gtggattagt gaacatggat taattatnttt 3076
Lys Asp Leu

gagataagag gaaagaaaag gagacatat atg gca tat att tgg tct tat ttg 3129
Met Ala Tyr Ile Trp Ser Tyr Leu
1010 1015

aaa agg tac ccc aat tgg tta tgg ctt gat tta cta gga gct atg ctt 3177
Lys Arg Tyr Pro Asn Trp Leu Trp Leu Asp Leu Leu Gly Ala Met Leu
1020 1025 1030

ttt gtg acg gtt atc cta gga atg ccc aca gcc tta gcg ggt atg att 3225
Phe Val Thr Val Ile Leu Gly Met Pro Thr Ala Leu Ala Gly Met Ile
1035 1040 1045

gat aat ggc gtt aca aaa ggt gat cgg act gga gtt tat ctg tgg acg 3273
Asp Asn Gly Val Thr Lys Gly Asp Arg Thr Gly Val Tyr Leu Trp Thr
1050 1055 1060

ttc atc atg ttt ata ttt gtt gta cta ggt att att ggg cgt att acg 3321
Phe Ile Met Phe Ile Phe Val Val Leu Gly Ile Ile Gly Arg Ile Thr
1065 1070 1075 1080

atg gct tac gca tct agt cgc tta acg aca aca atg att aga gat atg 3369
Met Ala Tyr Ala Ser Arg Leu Thr Thr Met Ile Arg Asp Met
1085 1090 1095

cgt aat gat atg tat gct aag ctt caa gaa tac tcc cat cat gaa tat 3417
Arg Asn Asp Met Tyr Ala Lys Leu Gln Glu Tyr Ser His His Glu Tyr
1100 1105 1110

gaa cag ata ggt gta tct tca cta gtg aca cgt atg aca agc gat act 3465
Glu Gln Ile Gly Val Ser Ser Leu Val Thr Arg Met Thr Ser Asp Thr
1115 1120 1125

ttt gtt ttg atg caa ttt gct gaa atg tct tta cgt tta ggc cta gta 3513
Phe Val Leu Met Gln Phe Ala Glu Met Ser Leu Arg Leu Gly Leu Val
1130 1135 1140

act cct atg gta atg att ttt agc gtg gtt atg ata cta att acg agt 3561
Thr Pro Met Val Met Ile Phe Ser Val Val Met Ile Leu Ile Thr Ser
1145 1150 1155 1160

cca tct ttg gct tgg ctt gta gcg gtt gcg atg cct ctt ttg gta gga 3609
Pro Ser Leu Ala Trp Leu Val Ala Val Ala Met Pro Leu Leu Val Gly
1165 1170 1175

gtc gtt tta tat gta gct ata aaa aca aaa cct tta tct gaa aga caa 3657
Val Val Leu Tyr Val Ala Ile Lys Thr Lys Pro Leu Ser Glu Arg Gln
1180 1185 1190

cag act atg ctt gat aaa atc aat caa tat gtt cgt gaa aat tta aca Gln Thr Met Leu Asp Lys Ile Asn Gln Tyr Val Arg Glu Asn Leu Thr 1195 1200 1205	3705
ggg tta cgc gtt gtt aga gcc ttt gca aga gag aat ttt caa tca caa Gly Leu Arg Val Val Arg Ala Phe Ala Arg Glu Asn Phe Gln Ser Gln 1210 1215 1220	3753
aaa ttt caa gtc gct aac caa cgt tac aca gat act tca act ggt ctt Lys Phe Gln Val Ala Asn Gln Arg Tyr Thr Asp Thr Ser Thr Gly Leu 1225 1230 1235 1240	3801
ttt aaa tta aca ggg cta aca gaa cca ctt ttc gtt caa att att att Phe Lys Leu Thr Gly Leu Thr Glu Pro Leu Phe Val Gln Ile Ile Ile 1245 1250 1255	3849
gca atg att gtg gct atc gtt tgg ttt gct ttg gat ccc tta caa aga Ala Met Ile Val Ala Ile Val Trp Phe Ala Leu Asp Pro Leu Gln Arg 1260 1265 1270	3897
ggt gct att aaa ata ggg gat tta gtt gct ttt atc gaa tat agc ttc Gly Ala Ile Lys Ile Gly Asp Leu Val Ala Phe Ile Glu Tyr Ser Phe 1275 1280 1285	3945
cat gct ctc ttt tca ttt ttg cta ttt gcc aat ctt ttt act atg tat His Ala Leu Phe Ser Phe Leu Leu Phe Ala Asn Leu Phe Thr Met Tyr 1290 1295 1300	3993
cct cgt atg gtg gta tca agc cat cgt att aga gag gtg atg gat atg Pro Arg Met Val Val Ser Ser His Arg Ile Arg Glu Val Met Asp Met 1305 1310 1315 1320	4041
cca atc tct atc aat cct aat gcc gaa ggt gtt acg gat acg aaa ctt Pro Ile Ser Ile Asn Pro Asn Ala Glu Gly Val Thr Asp Thr Lys Leu 1325 1330 1335	4089
aaa ggg cat tta gaa ttt gat aat gta aca ttc gct tat cca gga gaa Lys Gly His Leu Glu Phe Asp Asn Val Thr Phe Ala Tyr Pro Gly Glu 1340 1345 1350	4137
aca gag agt ccc gtt ttg cat gat att tct ttt aaa gct aag cct gga Thr Glu Ser Pro Val Leu His Asp Ile Ser Phe Lys Ala Lys Pro Gly 1355 1360 1365	4185
gaa aca att gct ttt att ggt tca aca ggt tca gga aaa tct tct ctt Glu Thr Ile Ala Phe Ile Gly Ser Thr Gly Ser Gly Lys Ser Ser Leu 1370 1375 1380	4233
ggt aat ttg att cca cgt ttt tat gat gtg aca ctt gga aaa atc tta Val Asn Leu Ile Pro Arg Phe Tyr Asp Val Thr Leu Gly Lys Ile Leu 1385 1390 1395 1400	4281
gta gat gga gtt gat gta aga gat tat aac ctt aaa tca ctt cgc caa Val Asp Gly Val Asp Val Arg Asp Tyr Asn Leu Lys Ser Leu Arg Gln 1405 1410 1415	4329

20 / 63

aag att gga ttt atc ccc caa aaa gct ctt tta ttt aca ggg aca ata Lys Ile Gly Phe Ile Pro Gln Lys Ala Leu Leu Phe Thr Gly Thr Ile 1420 1425 1430	4377
gga gag aat tta aaa tat gga aaa gct gat gct act att gat gat ctt Gly Glu Asn Leu Lys Tyr Gly Lys Ala Asp Ala Thr Ile Asp Asp Leu 1435 1440 1445	4425
aga caa gcg gtt gat att tct caa gct aaa gag ttt att gag agt cac Arg Gln Ala Val Asp Ile Ser Gln Ala Lys Glu Phe Ile Glu Ser His 1450 1455 1460	4473
caa gaa gcc ttt gaa acg cat tta gct gaa ggt ggg agc aat ctt tct Gln Glu Ala Phe Glu Thr His Leu Ala Glu Gly Gly Ser Asn Leu Ser 1465 1470 1475 1480	4521
ggg ggt caa aaa caa cgg tta tct att gct agg gct gtt gtt aaa gat Gly Gly Gln Lys Gln Arg Leu Ser Ile Ala Arg Ala Val Val Lys Asp 1485 1490 1495	4569
cca gat tta tat att ttt gat gat tca ttt tct gct ctc gat tat aag Pro Asp Leu Tyr Ile Phe Asp Asp Ser Phe Ser Ala Leu Asp Tyr Lys 1500 1505 1510	4617
aca gac gct act tta aga gcg cgt cta aaa gaa gta acc ggt gat tct Thr Asp Ala Thr Leu Arg Ala Arg Leu Lys Glu Val Thr Gly Asp Ser 1515 1520 1525	4665
aca gtt ttg ata gtt gct caa agg gtg ggt acg att atg gat gct gat Thr Val Leu Ile Val Ala Gln Arg Val Gly Thr Ile Met Asp Ala Asp 1530 1535 1540	4713
cag att att gtc ctt gat gaa ggc gaa att gtc ggt cgt ggt acc cac Gln Ile Ile Val Leu Asp Glu Gly Glu Ile Val Gly Arg Gly Thr His 1545 1550 1555 1560	4761
gct caa tta ata gaa aat aat gct att tat cgt gaa atc gct gag tca Ala Gln Leu Ile Glu Asn Asn Ala Ile Tyr Arg Glu Ile Ala Glu Ser 1565 1570 1575	4809
caa ctg aag aac caa aac tta tca gaa gga gag tgattgt atg aga aaa Gln Leu Lys Asn Gln Asn Leu Ser Glu Gly Glu Met Arg Lys 1580 1585 1590	4858
aaa tct gtt ttt ttg aga tta tgg tct tac cta act cgc tac aaa gct Lys Ser Val Phe Leu Arg Leu Trp Ser Tyr Leu Thr Arg Tyr Lys Ala 1595 1600 1605	4906
act ctt ttc tta gcg att ttt ttg aaa gtt tta tct agt ttt atg agt Thr Leu Phe Leu Ala Ile Phe Leu Lys Val Leu Ser Ser Phe Met Ser 1610 1615 1620	4954
gtt ctg gag cct ttt att tta ggg tta gcg ata aca gag ttg act gct Val Leu Glu Pro Phe Ile Leu Gly Leu Ala Ile Thr Glu Leu Thr Ala 1625 1630 1635	5002

aac ctt gtt gat atg gct aag gga gtt tct ggg gca gaa ttg aac gtt 5050
 Asn Leu Val Asp Met Ala Lys Gly Val Ser Gly Ala Glu Leu Asn Val
 1640 1645 1650

cct tat att gct ggt att ttg att att tat ttt ttc aga ggt gtt ttc 5098
 Pro Tyr Ile Ala Gly Ile Leu Ile Ile Tyr Phe Phe Arg Gly Val Phe
 1655 1660 1665 1670

tat gaa tta ggt tct tat ggc tca aat t 5126
 Tyr Glu Leu Gly Ser Tyr Gly Ser Asn
 1675

<210> 8
 <211> 229
 <212> PRT
 <213> Streptococcus

<400> 8
 Asn Phe Asp Ile Glu Thr Thr Thr Phe Glu Ala Met Lys Lys His Ala
 1 5 10 15
 Ser Leu Leu Glu Lys Ile Ser Val Glu Arg Ser Phe Ile Glu Phe Asp
 20 25 30
 Lys Leu Leu Leu Ala Pro Tyr Trp Arg Lys Gly Met Leu Ala Leu Ile
 35 40 45
 Asp Ser His Ala Phe Asn Tyr Leu Pro Cys Leu Lys Asn Arg Glu Leu
 50 55 60
 Gln Leu Ser Ala Phe Leu Ser Gln Leu Asp Lys Asp Phe Leu Phe Glu
 65 70 75 80
 Thr Ser Glu Gln Ala Trp Ala Ser Leu Ile Leu Ser Met Glu Val Glu
 85 90 95
 His Thr Lys Thr Phe Leu Lys Lys Trp Lys Thr Ser Thr His Phe Gln
 100 105 110
 Lys Asp Val Glu His Ile Val Asp Val Tyr Arg Ile Arg Glu Gln Met
 115 120 125
 Gly Leu Ala Lys Glu His Leu Tyr Arg Tyr Gly Lys Thr Ile Ile Lys
 130 135 140
 Gln Ala Glu Gly Ile Arg Lys Ala Arg Gly Leu Met Val Asp Phe Glu
 145 150 155 160
 Lys Ile Glu Gln Leu Asp Ser Glu Leu Ala Ile His Asp Arg His Glu
 165 170 175
 Ile Val Val Asn Gly Gly Thr Leu Ile Lys Lys Leu Gly Ile Lys Pro
 180 185 190
 Gly Pro Gln Met Gly Asp Ile Ile Ser Gln Ile Glu Leu Ala Ile Val
 195 200 205
 Leu Gly Gln Leu Ile Asn Glu Glu Ala Ile Leu His Phe Val Lys
 210 215 220
 Gln Tyr Leu Met Asp
 225

<210> 9
 <211> 622
 <212> PRT
 <213> Streptococcus

<400> 9
 Met Ser Asp Phe Leu Val Asp Gly Leu Thr Lys Ser Val Gly Asp Lys
 1 5 10 15
 Thr Val Phe Ser Asn Val Ser Phe Ile Ile His Ser Leu Asp Arg Ile
 20 25 30
 Gly Ile Ile Gly Val Asn Gly Thr Gly Lys Thr Thr Leu Leu Asp Val
 35 40 45
 Ile Ser Gly Glu Leu Gly Phe Asp Gly Asp Arg Ser Pro Phe Ser Ser
 50 55 60
 Ala Asn Asp Tyr Lys Ile Ala Tyr Leu Lys Gln Glu Pro Asp Phe Asp
 65 70 75 80
 Asp Ser Gln Thr Ile Leu Asp Thr Val Leu Ser Ser Asp Leu Arg Glu
 85 90 95
 Met Ala Leu Ile Lys Glu Tyr Glu Leu Leu Leu Asn His Tyr Glu Glu
 100 105 110
 Ser Lys Gln Ser Arg Leu Glu Lys Val Met Ala Glu Met Asp Ser Leu
 115 120 125
 Asp Ala Trp Ser Ile Glu Ser Glu Val Lys Thr Val Leu Ser Lys Leu
 130 135 140
 Gly Ile Thr Asp Leu Gln Leu Ser Val Gly Glu Leu Ser Gly Gly Leu
 145 150 155 160
 Arg Arg Arg Val Gln Leu Ala Gln Val Leu Leu Asn Asp Ala Asp Leu
 165 170 175
 Leu Leu Leu Asp Glu Pro Thr Asn His Leu Asp Ile Asp Thr Ile Ala
 180 185 190
 Trp Leu Thr Asn Phe Leu Lys Asn Ser Lys Lys Thr Val Leu Phe Ile
 195 200 205
 Thr His Asp Arg Tyr Phe Leu Asp Asn Val Ala Thr Arg Ile Phe Glu
 210 215 220
 Leu Asp Lys Ala Gln Ile Thr Glu Tyr Gln Gly Asn Tyr Gln Asp Tyr
 225 230 235 240
 Val Arg Leu Arg Ala Glu Gln Asp Glu Arg Asp Ala Ala Ser Leu His
 245 250 255
 Lys Lys Lys Gln Leu Tyr Lys Gln Glu Leu Ala Trp Met Arg Thr Gln
 260 265 270
 Pro Gln Ala Arg Ala Thr Lys Gln Gln Ala Arg Ile Asn Arg Phe Gln
 275 280 285
 Asn Leu Lys Asn Asp Leu His Gln Thr Ser Asp Thr Ser Asp Leu Glu
 290 295 300
 Met Thr Phe Glu Thr Ser Arg Ile Gly Lys Lys Val Ile Asn Phe Glu
 305 310 315 320
 Asn Val Ser Phe Ser Tyr Pro Asp Lys Ser Ile Leu Lys Asp Phe Asn
 325 330 335
 Leu Leu Ile Gln Asn Lys Asp Arg Ile Gly Ile Val Gly Asp Asn Gly
 340 345 350
 Val Gly Lys Ser Thr Leu Leu Asn Leu Ile Val Gln Asp Leu Gln Pro
 355 360 365
 Asp Ser Gly Asn Val Ser Ile Gly Glu Thr Ile Arg Val Gly Tyr Phe
 370 375 380
 Ser Gln Gln Leu His Asn Met Asp Gly Ser Lys Arg Val Ile Asn Tyr
 385 390 395 400
 Leu Gln Glu Val Ala Asp Glu Val Lys Thr Ser Val Gly Thr Thr Ser
 405 410 415
 Val Thr Glu Leu Leu Glu Gln Phe Leu Phe Pro Arg Ser Thr His Gly
 420 425 430

Thr Gln Ile Ala Lys Leu Ser Gly Gly Glu Lys Lys Arg Leu Tyr Leu
 435 440 445
 Leu Lys Ile Leu Ile Glu Lys Pro Asn Val Leu Leu Leu Asp Glu Pro
 450 455 460
 Thr Asn Asp Leu Asp Ile Ala Thr Leu Thr Val Leu Glu Asn Phe Leu
 465 470 475 480
 Gln Gly Phe Gly Gly Pro Val Ile Thr Val Ser His Asp Arg Tyr Phe
 485 490 495
 Leu Asp Lys Val Ala Asn Lys Ile Ile Ala Phe Glu Asp Asn Asp Ile
 500 505 510
 Arg Glu Phe Phe Gly Asn Tyr Thr Asp Tyr Leu Asp Glu Lys Ala Phe
 515 520 525
 Asn Glu Gln Asn Asn Glu Val Ile Ser Lys Lys Glu Ser Thr Lys Thr
 530 535 540
 Ser Arg Glu Lys Gln Ser Arg Lys Arg Met Ser Tyr Phe Glu Lys Gln
 545 550 555 560
 Glu Trp Ala Thr Ile Glu Asp Asp Ile Met Ile Leu Glu Asn Thr Ile
 565 570 575
 Thr Arg Ile Glu Asn Asp Met Gln Thr Cys Gly Ser Asp Phe Thr Arg
 580 585 590
 Leu Ser Asp Leu Gln Lys Glu Leu Asp Ala Lys Asn Glu Ala Leu Leu
 595 600 605
 Glu Lys Tyr Asp Arg Tyr Glu Tyr Leu Ser Glu Leu Asp Thr
 610 615 620

<210> 10
 <211> 157
 <212> PRT
 <213> Streptococcus

<400> 10
 Met Ile Ile Arg Pro Ile Ile Lys Asn Asp Asp Gln Ala Val Ala Gln
 1 5 10 15
 Leu Ile Arg Gln Ser Leu Arg Ala Tyr Asp Leu Asp Lys Pro Asp Thr
 20 25 30
 Ala Tyr Ser Asp Pro His Leu Asp His Leu Thr Ser Tyr Tyr Glu Lys
 35 40 45
 Ile Glu Lys Ser Gly Phe Phe Val Ile Glu Glu Arg Asp Glu Ile Ile
 50 55 60
 Gly Cys Gly Gly Phe Gly Pro Leu Lys Asn Leu Ile Ala Glu Met Gln
 65 70 75 80
 Lys Val Tyr Ile Ala Glu Arg Phe Arg Gly Lys Gly Leu Ala Thr Asp
 85 90 95
 Leu Val Lys Met Ile Glu Val Glu Ala Arg Lys Ile Gly Tyr Arg Gln
 100 105 110
 Leu Tyr Leu Glu Thr Ala Ser Thr Leu Ser Arg Ala Thr Ala Val Tyr
 115 120 125
 Lys His Met Gly Tyr Cys Ala Leu Ser Gln Pro Ile Ala Asn Asp Gln
 130 135 140
 Gly His Thr Ala Met Asp Ile Trp Met Ile Lys Asp Leu
 145 150 155

<210> 11
 <211> 579
 <212> PRT
 <213> Streptococcus

24 / 63

<400> 11
Met Ala Tyr Ile Trp Ser Tyr Leu Lys Arg Tyr Pro Asn Trp Leu Trp
1 5 10 15
Leu Asp Leu Leu Gly Ala Met Leu Phe Val Thr Val Ile Leu Gly Met
20 25 30
Pro Thr Ala Leu Ala Gly Met Ile Asp Asn Gly Val Thr Lys Gly Asp
35 40 45
Arg Thr Gly Val Tyr Leu Trp Thr Phe Ile Met Phe Ile Phe Val Val
50 55 60
Leu Gly Ile Ile Gly Arg Ile Thr Met Ala Tyr Ala Ser Ser Arg Leu
65 70 75 80
Thr Thr Thr Met Ile Arg Asp Met Arg Asn Asp Met Tyr Ala Lys Leu
85 90 95
Gln Glu Tyr Ser His His Glu Tyr Glu Gln Ile Gly Val Ser Ser Leu
100 105 110
Val Thr Arg Met Thr Ser Asp Thr Phe Val Leu Met Gln Phe Ala Glu
115 120 125
Met Ser Leu Arg Leu Gly Leu Val Thr Pro Met Val Met Ile Phe Ser
130 135 140
Val Val Met Ile Leu Ile Thr Ser Pro Ser Leu Ala Trp Leu Val Ala
145 150 155 160
Val Ala Met Pro Leu Leu Val Gly Val Val Leu Tyr Val Ala Ile Lys
165 170 175
Thr Lys Pro Leu Ser Glu Arg Gln Gln Thr Met Leu Asp Lys Ile Asn
180 185 190
Gln Tyr Val Arg Glu Asn Leu Thr Gly Leu Arg Val Val Arg Ala Phe
195 200 205
Ala Arg Glu Asn Phe Gln Ser Gln Lys Phe Gln Val Ala Asn Gln Arg
210 215 220
Tyr Thr Asp Thr Ser Thr Gly Leu Phe Lys Leu Thr Gly Leu Thr Glu
225 230 235 240
Pro Leu Phe Val Gln Ile Ile Ile Ala Met Ile Val Ala Ile Val Trp
245 250 255
Phe Ala Leu Asp Pro Leu Gln Arg Gly Ala Ile Lys Ile Gly Asp Leu
260 265 270
Val Ala Phe Ile Glu Tyr Ser Phe His Ala Leu Phe Ser Phe Leu Leu
275 280 285
Phe Ala Asn Leu Phe Thr Met Tyr Pro Arg Met Val Val Ser Ser His
290 295 300
Arg Ile Arg Glu Val Met Asp Met Pro Ile Ser Ile Asn Pro Asn Ala
305 310 315 320
Glu Gly Val Thr Asp Thr Lys Leu Lys Gly His Leu Glu Phe Asp Asn
325 330 335
Val Thr Phe Ala Tyr Pro Gly Glu Thr Glu Ser Pro Val Leu His Asp
340 345 350
Ile Ser Phe Lys Ala Lys Pro Gly Glu Thr Ile Ala Phe Ile Gly Ser
355 360 365
Thr Gly Ser Gly Lys Ser Ser Leu Val Asn Leu Ile Pro Arg Phe Tyr
370 375 380
Asp Val Thr Leu Gly Lys Ile Leu Val Asp Gly Val Asp Val Arg Asp
385 390 395 400
Tyr Asn Leu Lys Ser Leu Arg Gln Lys Ile Gly Phe Ile Pro Gln Lys
405 410 415
Ala Leu Leu Phe Thr Gly Thr Ile Gly Glu Asn Leu Lys Tyr Gly Lys
420 425 430

Ala Asp Ala Thr Ile Asp Asp Leu Arg Gln Ala Val Asp Ile Ser Gln
 435 440 445
 Ala Lys Glu Phe Ile Glu Ser His Gln Glu Ala Phe Glu Thr His Leu
 450 455 460
 Ala Glu Gly Gly Ser Asn Leu Ser Gly Gly Gln Lys Gln Arg Leu Ser
 465 470 475 480
 Ile Ala Arg Ala Val Val Lys Asp Pro Asp Leu Tyr Ile Phe Asp Asp
 485 490 495
 Ser Phe Ser Ala Leu Asp Tyr Lys Thr Asp Ala Thr Leu Arg Ala Arg
 500 505 510
 Leu Lys Glu Val Thr Gly Asp Ser Thr Val Leu Ile Val Ala Gln Arg
 515 520 525
 Val Gly Thr Ile Met Asp Ala Asp Gln Ile Ile Val Leu Asp Glu Gly
 530 535 540
 Glu Ile Val Gly Arg Gly Thr His Ala Gln Leu Ile Glu Asn Asn Ala
 545 550 555 560
 Ile Tyr Arg Glu Ile Ala Glu Ser Gln Leu Lys Asn Gln Asn Leu Ser
 565 570 575
 Glu Gly Glu

<210> 12
 <211> 92
 <212> PRT
 <213> Streptococcus

<400> 12
 Met Arg Lys Lys Ser Val Phe Leu Arg Leu Trp Ser Tyr Leu Thr Arg
 1 5 10 15
 Tyr Lys Ala Thr Leu Phe Leu Ala Ile Phe Leu Lys Val Leu Ser Ser
 20 25 30
 Phe Met Ser Val Leu Glu Pro Phe Ile Leu Gly Leu Ala Ile Thr Glu
 35 40 45
 Leu Thr Ala Asn Leu Val Asp Met Ala Lys Gly Val Ser Gly Ala Glu
 50 55 60
 Leu Asn Val Pro Tyr Ile Ala Gly Ile Leu Ile Ile Tyr Phe Phe Arg
 65 70 75 80
 Gly Val Phe Tyr Glu Leu Gly Ser Tyr Gly Ser Asn
 85 90

<210> 13
 <211> 5215
 <212> DNA
 <213> Streptococcus

<220>
 <221> CDS
 <222> (3)...(122)

 <221> CDS
 <222> (133)...(2511)

 <221> CDS
 <222> (367)...(2511)

 <221> CDS

```

<222> (2946)...(2716)
<223> of complementary strand

<221> CDS
<222> (3252)...(2995)
<223> of complementary strand

<221> CDS
<222> (3676)...(3299)
<223> of complementary strand

<221> CDS
<222> (4124)...(3837)
<223> of complementary strand

<221> CDS
<222> (5214)...(4351)
<223> of complementary strand

<400> 13
aa ttt gga agt gct cta tca aca gtt gaa gta aag gag att att agt      47
  Phe Gly Ser Ala Leu Ser Thr Val Glu Val Lys Glu Ile Ile Ser
    1             5             10             15

gaa gaa aac ata tgg tta tat cgg ctc agt tgc tgc cat ttt act agc      95
Glu Glu Asn Ile Trp Leu Tyr Arg Leu Ser Cys Cys His Phe Thr Ser
                20             25             30

tac tca tat tgg aag tta cca act tgg taagcatcat atg ggt cta gca      144
Tyr Ser Tyr Trp Lys Leu Pro Thr Trp           Met Gly Leu Ala
                35             40

aca aag gac aat cag att gcc tat att gat gac agc aaa ggt aag gca      192
Thr Lys Asp Asn Gln Ile Ala Tyr Ile Asp Asp Ser Lys Gly Lys Ala
    45             50             55             60

aaa gcc cct aaa aca aac aaa acg atg gat caa atc agt gct gaa gaa      240
Lys Ala Pro Lys Thr Asn Lys Thr Met Asp Gln Ile Ser Ala Glu Glu
                65             70             75

ggc atc tct gct gaa cag atc gta gtc aaa att act gac caa ggc tat      288
Gly Ile Ser Ala Glu Gln Ile Val Val Lys Ile Thr Asp Gln Gly Tyr
                80             85             90

gtg acc tca cac ggt gac cat tat cat ttt tac aat ggg aaa gtt cct      336
Val Thr Ser His Gly Asp His Tyr His Phe Tyr Asn Gly Lys Val Pro
                95             100            105

tat gat gcg att att agt gaa gag ttg ttg atg acg gat cct aat tac      384
Tyr Asp Ala Ile Ile Ser Glu Glu Leu Leu Met Thr Asp Pro Asn Tyr
    110             115             120

cgt ttt aaa caa tca gac gtt atc aat gaa atc tta gac ggt tac gtt      432
Arg Phe Lys Gln Ser Asp Val Ile Asn Glu Ile Leu Asp Gly Tyr Val
    125             130             135             140

```

27 / 63

att	aaa	gtc	aat	ggc	aac	tat	tat	gtt	tac	ctc	aag	cca	ggt	agt	aag	480
Ile	Lys	Val	Asn	Gly	Asn	Tyr	Tyr	Val	Tyr	Leu	Lys	Pro	Gly	Ser	Lys	
				145					150						155	
cgc	aaa	aac	att	cga	acc	aaa	caa	caa	att	gct	gag	caa	gta	gcc	aaa	528
Arg	Lys	Asn	Ile	Arg	Thr	Lys	Gln	Gln	Ile	Ala	Glu	Gln	Val	Ala	Lys	
			160					165					170			
gga	act	aaa	gaa	gct	aaa	gaa	aaa	ggt	tta	gct	caa	gtg	gcc	cat	ctc	576
Gly	Thr	Lys	Glu	Ala	Lys	Glu	Lys	Gly	Leu	Ala	Gln	Val	Ala	His	Leu	
		175					180						185			
agt	aaa	gaa	gaa	gtt	gcg	gca	gtc	aat	gaa	gca	aaa	aga	caa	gga	cgc	624
Ser	Lys	Glu	Glu	Val	Ala	Ala	Val	Asn	Glu	Ala	Lys	Arg	Gln	Gly	Arg	
	190					195					200					
tat	act	aca	gac	gat	ggc	tat	att	ttt	agt	ccg	aca	gat	atc	att	gat	672
Tyr	Thr	Thr	Asp	Asp	Gly	Tyr	Ile	Phe	Ser	Pro	Thr	Asp	Ile	Ile	Asp	
	205				210					215					220	
gat	tta	gga	gat	gct	tat	tta	gta	cct	cat	ggt	aat	cac	tat	cat	tat	720
Asp	Leu	Gly	Asp	Ala	Tyr	Leu	Val	Pro	His	Gly	Asn	His	Tyr	His	Tyr	
				225					230					235		
att	cct	aaa	aag	gat	ttg	tct	cca	agt	gag	cta	gct	gct	gca	caa	gcc	768
Ile	Pro	Lys	Lys	Asp	Leu	Ser	Pro	Ser	Glu	Leu	Ala	Ala	Ala	Gln	Ala	
			240					245					250			
tac	tgg	agt	caa	aaa	caa	ggt	cga	ggt	gct	aga	ccg	tct	gat	tac	cgc	816
Tyr	Trp	Ser	Gln	Lys	Gln	Gly	Arg	Gly	Ala	Arg	Pro	Ser	Asp	Tyr	Arg	
		255					260					265				
ccg	aca	cca	gcc	cca	ggt	cgt	agg	aaa	gcc	cca	att	cct	gat	gtg	acg	864
Pro	Thr	Pro	Ala	Pro	Gly	Arg	Arg	Lys	Ala	Pro	Ile	Pro	Asp	Val	Thr	
	270					275					280					
cct	aac	cct	gga	caa	ggt	cat	cag	cca	gat	aac	ggt	ggc	tat	cat	cca	912
Pro	Asn	Pro	Gly	Gln	Gly	His	Gln	Pro	Asp	Asn	Gly	Gly	Tyr	His	Pro	
	285				290					295					300	
gcg	cct	cct	agg	cca	aat	gat	gcg	tca	caa	aac	aaa	cac	caa	aga	gat	960
Ala	Pro	Pro	Arg	Pro	Asn	Asp	Ala	Ser	Gln	Asn	Lys	His	Gln	Arg	Asp	
				305					310					315		
gag	ttt	aaa	gga	aaa	acc	ttt	aag	gaa	ctt	tta	gat	caa	cta	cac	cg	1008
Glu	Phe	Lys	Gly	Lys	Thr	Phe	Lys	Glu	Leu	Leu	Asp	Gln	Leu	His	Arg	
			320					325					330			
ctt	gat	ttg	aaa	tac	cgt	cat	gtg	gaa	gaa	gat	ggg	ttg	att	ttt	gaa	1056
Leu	Asp	Leu	Lys	Tyr	Arg	His	Val	Glu	Glu	Asp	Gly	Leu	Ile	Phe	Glu	
		335					340					345				
ccg	act	caa	gtg	atc	aaa	tca	aac	gct	ttt	ggg	tat	gtg	gtg	cct	cat	1104
Pro	Thr	Gln	Val	Ile	Lys	Ser	Asn	Ala	Phe	Gly	Tyr	Val	Val	Pro	His	
	350					355					360					

28 / 63

gga gat cat tat cat att atc cca aga agt cag tta tca cct ctt gaa	1152
Gly Asp His Tyr His Ile Ile Pro Arg Ser Gln Leu Ser Pro Leu Glu	
365 370 375 380	
atg gaa tta gca gat cga tac tta gct ggc caa act gag gac aat gac	1200
Met Glu Leu Ala Asp Arg Tyr Leu Ala Gly Gln Thr Glu Asp Asn Asp	
385 390 395	
tca ggt tca gag cac tca aaa cca tca gat aaa gaa gtg aca cat acc	1248
Ser Gly Ser Glu His Ser Lys Pro Ser Asp Lys Glu Val Thr His Thr	
400 405 410	
ttt ctt ggt cat cgc atc aaa gct tac gga aaa ggc tta gat ggt aaa	1296
Phe Leu Gly His Arg Ile Lys Ala Tyr Gly Lys Gly Leu Asp Gly Lys	
415 420 425	
cca tat gat acg agt gat gct tat gtt ttt agt aaa gaa tcc att cat	1344
Pro Tyr Asp Thr Ser Asp Ala Tyr Val Phe Ser Lys Glu Ser Ile His	
430 435 440	
tca gtg gat aaa tca gga gtt aca gct aaa cac gga gat cat ttc cac	1392
Ser Val Asp Lys Ser Gly Val Thr Ala Lys His Gly Asp His Phe His	
445 450 455 460	
tat ata gga ttt gga gaa ctt gaa caa tat gag ttg gat gag gtc gct	1440
Tyr Ile Gly Phe Gly Glu Leu Glu Gln Tyr Glu Leu Asp Glu Val Ala	
465 470 475	
aac tgg gtg aaa gca aaa ggt caa gct gat gag ctt gct gct gct ttg	1488
Asn Trp Val Lys Ala Lys Gly Gln Ala Asp Glu Leu Ala Ala Ala Leu	
480 485 490	
gat cag gaa caa ggc aaa gaa aaa cca ctc ttt gac act aaa aaa gtg	1536
Asp Gln Glu Gln Gly Lys Glu Lys Pro Leu Phe Asp Thr Lys Lys Val	
495 500 505	
agt cgc aaa gta aca aaa gat ggt aaa gtg ggc tat atg atg cca aaa	1584
Ser Arg Lys Val Thr Lys Asp Gly Lys Val Gly Tyr Met Met Pro Lys	
510 515 520	
gat ggt aag gac tat ttc tat gct cgt gat caa ctt gat ttg act cag	1632
Asp Gly Lys Asp Tyr Phe Tyr Ala Arg Asp Gln Leu Asp Leu Thr Gln	
525 530 535 540	
att gcc ttt gcc gaa caa gaa cta atg ctt aaa gat aag aag cat tac	1680
Ile Ala Phe Ala Glu Gln Glu Leu Met Leu Lys Asp Lys Lys His Tyr	
545 550 555	
cgt tat gac att gtt gac aca ggt att gag cca cga ctt gct gta gat	1728
Arg Tyr Asp Ile Val Asp Thr Gly Ile Glu Pro Arg Leu Ala Val Asp	
560 565 570	
gtg tca agt ctg ccg atg cat gct ggt aat gct act tac gat act gga	1776
Val Ser Ser Leu Pro Met His Ala Gly Asn Ala Thr Tyr Asp Thr Gly	
575 580 585	

agt tcg ttt gtt atc cca cat att gat cat atc cat gtc gtt ccg tat	1824
Ser Ser Phe Val Ile Pro His Ile Asp His Ile His Val Val Pro Tyr	
590 595 600	
tca tgg ttg acg cgc gat cag att gca aca gtc aag tat gtg atg caa	1872
Ser Trp Leu Thr Arg Asp Gln Ile Ala Thr Val Lys Tyr Val Met Gln	
605 610 615 620	
cac ccc gaa gtt cgt ccg gat gta tgg tct aag cca ggg cat gaa gag	1920
His Pro Glu Val Arg Pro Asp Val Trp Ser Lys Pro Gly His Glu Glu	
625 630 635	
tca ggt tcg gtc att cca aat gtt acg cct ctt gat aaa cgt gct ggt	1968
Ser Gly Ser Val Ile Pro Asn Val Thr Pro Leu Asp Lys Arg Ala Gly	
640 645 650	
atg cca aac tgg caa att atc cat tct gct gaa gaa gtt caa aaa gcc	2016
Met Pro Asn Trp Gln Ile Ile His Ser Ala Glu Glu Val Gln Lys Ala	
655 660 665	
cta gca gaa ggt cgt ttt gca aca cca gac ggc tat att ttc gat cca	2064
Leu Ala Glu Gly Arg Phe Ala Thr Pro Asp Gly Tyr Ile Phe Asp Pro	
670 675 680	
cga gat gtt ttg gcc aaa gaa act ttt gta tgg aaa gat ggc tcc ttt	2112
Arg Asp Val Leu Ala Lys Glu Thr Phe Val Trp Lys Asp Gly Ser Phe	
685 690 695 700	
agc atc cca aga gca gat ggc agt tca ttg aga acc att aat aaa tct	2160
Ser Ile Pro Arg Ala Asp Gly Ser Ser Leu Arg Thr Ile Asn Lys Ser	
705 710 715	
gat cta tcc caa gct gag tgg caa caa gct caa gag tta ttg gca aag	2208
Asp Leu Ser Gln Ala Glu Trp Gln Gln Ala Gln Glu Leu Leu Ala Lys	
720 725 730	
aaa aat act ggt gat gct act gat acg gat aaa ccc aaa gaa aag caa	2256
Lys Asn Thr Gly Asp Ala Thr Asp Thr Asp Lys Pro Lys Glu Lys Gln	
735 740 745	
cag gca gat aag agc aat gaa aac caa cag cca agt gaa gcc agt aaa	2304
Gln Ala Asp Lys Ser Asn Glu Asn Gln Gln Pro Ser Glu Ala Ser Lys	
750 755 760	
gaa gaa aaa gaa tca gat gac ttt ata gac agt tta cca gac tat ggt	2352
Glu Glu Lys Glu Ser Asp Asp Phe Ile Asp Ser Leu Pro Asp Tyr Gly	
765 770 775 780	
cta gat aga gca acc cta gaa gat cat atc aat caa tta gca caa aaa	2400
Leu Asp Arg Ala Thr Leu Glu Asp His Ile Asn Gln Leu Ala Gln Lys	
785 790 795	
gct aat atc gat cct aag tat ctc att ttc caa cca gaa ggt gtc caa	2448
Ala Asn Ile Asp Pro Lys Tyr Leu Ile Phe Gln Pro Glu Gly Val Gln	
800 805 810	

ttt tat aat aaa aat ggt gaa ttg gta act tat gat atc aag aca ctt	2496
Phe Tyr Asn Lys Asn Gly Glu Leu Val Thr Tyr Asp Ile Lys Thr Leu	
815	820
	825

caa caa ata aac cct taacccaaaag aagatctcat tgtaaagca ctgctttgtc	2551
Gln Gln Ile Asn Pro	
830	

aaagcaagtt	acggtgattt	tgaagtcatt	ctatgtaacg	agtagtgata	aaagttggat	2611
aatagcgggt	ttcttttgca	aagaaaagg	atccatgta	gaatagtaaa	aaaagaggag	2671
gattcttggg	ctaagtcaaa	ataagtagac	agaaaactgt	gttattttattg	cggt	2726
taaaataaatt	ttcttctttc	tgattagggg	ttagtcctag	attagccgta	tgtgggttgt	2786
aattgttata	aaaattctca	atgtattcaa	agcagctcaa	ttgaacctgt	ttgatatttt	2846
gataatggtt	tcggttgatt	tgtctatgct	ttaaatactt	gaaaaatgct	tcagttacgg	2906
cattatcata	aggatatcca	ggattagaaa	aagaatgcat	gatattggca	ctgcacccta	2966
atagtggagc	gcaagaaaaa	cacttttagg	caatcagtt	ttctgtactg	tacaggcgac	3025
tggtcggtta	atctctgttg	aattctagtt	tcattataaa	atgtaatgta	atttttaaca	3085
atatttggtt	tactatcttt	gttgtatttt	ctcctattat	ggaaataaaa	ggtttcagtc	3145
tttaggacgg	tgtgaaacca	ttcaatacag	gcattatctg	cagggtgttc	ttttcgagac	3205
attgagcggg	taatgtcttt	ttccgtgcaa	gcctggtagt	aagccataga	agtatacact	3265
gagccttggg	cactgtgtaa	gattgctcct	ttatttaggca	att	ttaactgatt	3319
aagggtgtct	agtacaaaat	ccgtgtcctg	acaatctgag	atagtgtaag	ctataatttc	3379
tcggttatag	agattcataa	ttgatgagag	atacaattta	cagttaccga	aatataggta	3439
ggtaaatatc	gttacgagct	tttccttagg	cttatcggca	tggaaatccc	gactcaattt	3499
attatctggt	aaataataag	ctttacccaa	attgggaact	ttcttggtac	gtgtccgaca	3559
aagccagcca	ttattttttca	tgatacgata	gactttcttt	gtattaacag	tcaatccgtg	3619
gattttttttg	agcaatcgtg	taatggtagc	atagccataa	ataaagtgat	tctccataca	3679
gagctgttca	attaattcaa	taaggctatc	tttttttggc	gcttctcata	ctcctttttc	3739
caacggtaat	aggtcgaccg	cttgacctta	aaacagtcta	gaatgaaaac	tatcgggtag	3799
ttgtttttat	agtcttccac	aagcttgata	agacttacttt	atcgcatt	tccttatcaa	3857
gctcagatac	tttttttaaga	ggcaacctg	taattgtaat	tgttccactt	cagacagatg	3917
ttccaagcct	ttaccgtagg	tatattgctt	gccaacacct	tgatgaaaac	gataaagctc	3977
ctcgttttcg	taccattttca	tccaagtata	gatttgacta	ttattttttga	tgccataaag	4037
ctccataata	actctgttag	acttgctctg	tttcttcata	tcgatgcaag	ccagcttagt	4097
ttcccatgaa	tatgcttttt	taaccataat	aaaacattcc	tgtttctagt	ttactaaatt	4157
tcaacaggag	tgtttttctt	ttgtctcatt	ttagggattc	agtgcctatt	gttgtcatca	4217
attattttttc	taaattcccc	ggacttaaat	tgtgacctt	ggtcgggaatg	aaagagaagt	4277
gttcttccaa	tctttctttt	attaagttaa	aaggcaaac	ttttctgtac	aacatttata	4337
aagtgttttt	ctaggcaatta	atc ttttagt	cat tgggtgtt	gg tagttgagac		4391
taccatgaat	gcggtggtaa	ttccaccaat	gaacatagtc	tttagtctta	agagctagtt	4451
cttccagcaa	ttgaaagggt	tcttgataaa	caaattcaat	tttgaaagca	cgatacgtac	4511
tttcagctac	ggcattgtca	taaggataac	cagcctgact	aagcgaactg	gtgattccaa	4571
aggcttccaa	tatttcatca	attaactgat	tatcaaacct	tttgccacga	tctgaatgga	4631
acatcttgac	tttgggtcagg	gcgtaaggga	tgctttgtat	ggcttgctta	acgagttcag	4691
cggtcttgtg	ccaaccaaga	gacaggccga	tgatttcacg	gttgtatagg	tcaatgatga	4751
ggcaaacata	agcccaacga	ttgcctacac	gaacataggt	taagtcagtg	actaaggctt	4811
gtagtggctc	ttcttgctta	aattgcctgt	ctaagtgggt	gggaataggg	gcttctattc	4871
tgctctaga	atgtgggttg	aagggtggctt	tctgataaac	agaaaccaa	ttgagtcgct	4931
tcataatgcg	tcgaatccga	cgacgtgaaa	gtgtgatacc	ttcgttattc	aagcatattt	4991
tgattttttct	ggatccgtat	ctagactcgc	tatcgagaaa	aattctttta	atagtttctt	5051
caaactccgt	ttcagatact	gactccacgg	cttgatagta	ataacttgag	tgtggcatat	5111
tcagccagcg	acacatcttt	gaaatgctgt	atctatcctt	attagcagtg	attatttccc	5171
ttttttgtcc	ataatcccg	ctgcttgctt	taggatattc	aatt		5215

<210> 14
 <211> 40
 <212> PRT

<213> Streptococcus

<400> 14

Phe Gly Ser Ala Leu Ser Thr Val Glu Val Lys Glu Ile Ile Ser Glu
 1 5 10 15
 Glu Asn Ile Trp Leu Tyr Arg Leu Ser Cys Cys His Phe Thr Ser Tyr
 20 25 30
 Ser Tyr Trp Lys Leu Pro Thr Trp
 35 40

<210> 15

<211> 793

<212> PRT

<213> Streptococcus

<400> 15

Met Gly Leu Ala Thr Lys Asp Asn Gln Ile Ala Tyr Ile Asp Asp Ser
 1 5 10 15
 Lys Gly Lys Ala Lys Ala Pro Lys Thr Asn Lys Thr Met Asp Gln Ile
 20 25 30
 Ser Ala Glu Glu Gly Ile Ser Ala Glu Gln Ile Val Val Lys Ile Thr
 35 40 45
 Asp Gln Gly Tyr Val Thr Ser His Gly Asp His Tyr His Phe Tyr Asn
 50 55 60
 Gly Lys Val Pro Tyr Asp Ala Ile Ile Ser Glu Glu Leu Leu Met Thr
 65 70 75 80
 Asp Pro Asn Tyr Arg Phe Lys Gln Ser Asp Val Ile Asn Glu Ile Leu
 85 90 95
 Asp Gly Tyr Val Ile Lys Val Asn Gly Asn Tyr Tyr Val Tyr Leu Lys
 100 105 110
 Pro Gly Ser Lys Arg Lys Asn Ile Arg Thr Lys Gln Gln Ile Ala Glu
 115 120 125
 Gln Val Ala Lys Gly Thr Lys Glu Ala Lys Glu Lys Gly Leu Ala Gln
 130 135 140
 Val Ala His Leu Ser Lys Glu Glu Val Ala Ala Val Asn Glu Ala Lys
 145 150 155 160
 Arg Gln Gly Arg Tyr Thr Thr Asp Asp Gly Tyr Ile Phe Ser Pro Thr
 165 170 175
 Asp Ile Ile Asp Asp Leu Gly Asp Ala Tyr Leu Val Pro His Gly Asn
 180 185 190
 His Tyr His Tyr Ile Pro Lys Lys Asp Leu Ser Pro Ser Glu Leu Ala
 195 200 205
 Ala Ala Gln Ala Tyr Trp Ser Gln Lys Gln Gly Arg Gly Ala Arg Pro
 210 215 220
 Ser Asp Tyr Arg Pro Thr Pro Ala Pro Gly Arg Arg Lys Ala Pro Ile
 225 230 235 240
 Pro Asp Val Thr Pro Asn Pro Gly Gln Gly His Gln Pro Asp Asn Gly
 245 250 255
 Gly Tyr His Pro Ala Pro Pro Arg Pro Asn Asp Ala Ser Gln Asn Lys
 260 265 270
 His Gln Arg Asp Glu Phe Lys Gly Lys Thr Phe Lys Glu Leu Leu Asp
 275 280 285
 Gln Leu His Arg Leu Asp Leu Lys Tyr Arg His Val Glu Glu Asp Gly
 290 295 300
 Leu Ile Phe Glu Pro Thr Gln Val Ile Lys Ser Asn Ala Phe Gly Tyr
 305 310 315 320

Val Val Pro His Gly Asp His Tyr His Ile Ile Pro Arg Ser Gln Leu
 325 330 335
 Ser Pro Leu Glu Met Glu Leu Ala Asp Arg Tyr Leu Ala Gly Gln Thr
 340 345 350
 Glu Asp Asn Asp Ser Gly Ser Glu His Ser Lys Pro Ser Asp Lys Glu
 355 360 365
 Val Thr His Thr Phe Leu Gly His Arg Ile Lys Ala Tyr Gly Lys Gly
 370 375 380
 Leu Asp Gly Lys Pro Tyr Asp Thr Ser Asp Ala Tyr Val Phe Ser Lys
 385 390 395 400
 Glu Ser Ile His Ser Val Asp Lys Ser Gly Val Thr Ala Lys His Gly
 405 410 415
 Asp His Phe His Tyr Ile Gly Phe Gly Glu Leu Glu Gln Tyr Glu Leu
 420 425 430
 Asp Glu Val Ala Asn Trp Val Lys Ala Lys Gly Gln Ala Asp Glu Leu
 435 440 445
 Ala Ala Ala Leu Asp Gln Glu Glu Gly Lys Glu Lys Pro Leu Phe Asp
 450 455 460
 Thr Lys Lys Val Ser Arg Lys Val Thr Lys Asp Gly Lys Val Gly Tyr
 465 470 475 480
 Met Met Pro Lys Asp Gly Lys Asp Tyr Phe Tyr Ala Arg Asp Gln Leu
 485 490 495
 Asp Leu Thr Gln Ile Ala Phe Ala Glu Gln Glu Leu Met Leu Lys Asp
 500 505 510
 Lys Lys His Tyr Arg Tyr Asp Ile Val Asp Thr Gly Ile Glu Pro Arg
 515 520 525
 Leu Ala Val Asp Val Ser Ser Leu Pro Met His Ala Gly Asn Ala Thr
 530 535 540
 Tyr Asp Thr Gly Ser Ser Phe Val Ile Pro His Ile Asp His Ile His
 545 550 555 560
 Val Val Pro Tyr Ser Trp Leu Thr Arg Asp Gln Ile Ala Thr Val Lys
 565 570 575
 Tyr Val Met Gln His Pro Glu Val Arg Pro Asp Val Trp Ser Lys Pro
 580 585 590
 Gly His Glu Glu Ser Gly Ser Val Ile Pro Asn Val Thr Pro Leu Asp
 595 600 605
 Lys Arg Ala Gly Met Pro Asn Trp Gln Ile Ile His Ser Ala Glu Glu
 610 615 620
 Val Gln Lys Ala Leu Ala Glu Gly Arg Phe Ala Thr Pro Asp Gly Tyr
 625 630 635 640
 Ile Phe Asp Pro Arg Asp Val Leu Ala Lys Glu Thr Phe Val Trp Lys
 645 650 655
 Asp Gly Ser Phe Ser Ile Pro Arg Ala Asp Gly Ser Ser Leu Arg Thr
 660 665 670
 Ile Asn Lys Ser Asp Leu Ser Gln Ala Glu Trp Gln Gln Ala Gln Glu
 675 680 685
 Leu Leu Ala Lys Lys Asn Thr Gly Asp Ala Thr Asp Thr Asp Lys Pro
 690 695 700
 Lys Glu Lys Gln Gln Ala Asp Lys Ser Asn Glu Asn Gln Gln Pro Ser
 705 710 715 720
 Glu Ala Ser Lys Glu Glu Lys Glu Ser Asp Asp Phe Ile Asp Ser Leu
 725 730 735
 Pro Asp Tyr Gly Leu Asp Arg Ala Thr Leu Glu Asp His Ile Asn Gln
 740 745 750
 Leu Ala Gln Lys Ala Asn Ile Asp Pro Lys Tyr Leu Ile Phe Gln Pro
 755 760 765

Glu Gly Val Gln Phe Tyr Asn Lys Asn Gly Glu Leu Val Thr Tyr Asp
 770 775 780
 Ile Lys Thr Leu Gln Gln Ile Asn Pro
 785 790

<210> 16
 <211> 715
 <212> PRT
 <213> Streptococcus

<400> 16
 Met Thr Asp Pro Asn Tyr Arg Phe Lys Gln Ser Asp Val Ile Asn Glu
 1 5 10 15
 Ile Leu Asp Gly Tyr Val Ile Lys Val Asn Gly Asn Tyr Tyr Val Tyr
 20 25 30
 Leu Lys Pro Gly Ser Lys Arg Lys Asn Ile Arg Thr Lys Gln Gln Ile
 35 40 45
 Ala Glu Gln Val Ala Lys Gly Thr Lys Glu Ala Lys Glu Lys Gly Leu
 50 55 60
 Ala Gln Val Ala His Leu Ser Lys Glu Glu Val Ala Ala Val Asn Glu
 65 70 75 80
 Ala Lys Arg Gln Gly Arg Tyr Thr Thr Asp Asp Gly Tyr Ile Phe Ser
 85 90 95
 Pro Thr Asp Ile Ile Asp Asp Leu Gly Asp Ala Tyr Leu Val Pro His
 100 105 110
 Gly Asn His Tyr His Tyr Ile Pro Lys Lys Asp Leu Ser Pro Ser Glu
 115 120 125
 Leu Ala Ala Ala Gln Ala Tyr Trp Ser Gln Lys Gln Gly Arg Gly Ala
 130 135 140
 Arg Pro Ser Asp Tyr Arg Pro Thr Pro Ala Pro Gly Arg Arg Lys Ala
 145 150 155 160
 Pro Ile Pro Asp Val Thr Pro Asn Pro Gly Gln Gly His Gln Pro Asp
 165 170 175
 Asn Gly Gly Tyr His Pro Ala Pro Pro Arg Pro Asn Asp Ala Ser Gln
 180 185 190
 Asn Lys His Gln Arg Asp Glu Phe Lys Gly Lys Thr Phe Lys Glu Leu
 195 200 205
 Leu Asp Gln Leu His Arg Leu Asp Leu Lys Tyr Arg His Val Glu Glu
 210 215 220
 Asp Gly Leu Ile Phe Glu Pro Thr Gln Val Ile Lys Ser Asn Ala Phe
 225 230 235 240
 Gly Tyr Val Val Pro His Gly Asp His Tyr His Ile Ile Pro Arg Ser
 245 250 255
 Gln Leu Ser Pro Leu Glu Met Glu Leu Ala Asp Arg Tyr Leu Ala Gly
 260 265 270
 Gln Thr Glu Asp Asn Asp Ser Gly Ser Glu His Ser Lys Pro Ser Asp
 275 280 285
 Lys Glu Val Thr His Thr Phe Leu Gly His Arg Ile Lys Ala Tyr Gly
 290 295 300
 Lys Gly Leu Asp Gly Lys Pro Tyr Asp Thr Ser Asp Ala Tyr Val Phe
 305 310 315 320
 Ser Lys Glu Ser Ile His Ser Val Asp Lys Ser Gly Val Thr Ala Lys
 325 330 335
 His Gly Asp His Phe His Tyr Ile Gly Phe Gly Glu Leu Glu Gln Tyr
 340 345 350

Glu Leu Asp Glu Val Ala Asn Trp Val Lys Ala Lys Gly Gln Ala Asp
 355 360 365
 Glu Leu Ala Ala Ala Leu Asp Gln Glu Gln Gly Lys Glu Lys Pro Leu
 370 375 380
 Phe Asp Thr Lys Lys Val Ser Arg Lys Val Thr Lys Asp Gly Lys Val
 385 390 395 400
 Gly Tyr Met Met Pro Lys Asp Gly Lys Asp Tyr Phe Tyr Ala Arg Asp
 405 410 415
 Gln Leu Asp Leu Thr Gln Ile Ala Phe Ala Glu Gln Glu Leu Met Leu
 420 425 430
 Lys Asp Lys Lys His Tyr Arg Tyr Asp Ile Val Asp Thr Gly Ile Glu
 435 440 445
 Pro Arg Leu Ala Val Asp Val Ser Ser Leu Pro Met His Ala Gly Asn
 450 455 460
 Ala Thr Tyr Asp Thr Gly Ser Ser Phe Val Ile Pro His Ile Asp His
 465 470 475 480
 Ile His Val Val Pro Tyr Ser Trp Leu Thr Arg Asp Gln Ile Ala Thr
 485 490 495
 Val Lys Tyr Val Met Gln His Pro Glu Val Arg Pro Asp Val Trp Ser
 500 505 510
 Lys Pro Gly His Glu Glu Ser Gly Ser Val Ile Pro Asn Val Thr Pro
 515 520 525
 Leu Asp Lys Arg Ala Gly Met Pro Asn Trp Gln Ile His Ser Ala
 530 535 540
 Glu Glu Val Gln Lys Ala Leu Ala Glu Gly Arg Phe Ala Thr Pro Asp
 545 550 555 560
 Gly Tyr Ile Phe Asp Pro Arg Asp Val Leu Ala Lys Glu Thr Phe Val
 565 570 575
 Trp Lys Asp Gly Ser Phe Ser Ile Pro Arg Ala Asp Gly Ser Ser Leu
 580 585 590
 Arg Thr Ile Asn Lys Ser Asp Leu Ser Gln Ala Glu Trp Gln Gln Ala
 595 600 605
 Gln Glu Leu Leu Ala Lys Lys Asn Thr Gly Asp Ala Thr Asp Thr Asp
 610 615 620
 Lys Pro Lys Glu Lys Gln Gln Ala Asp Lys Ser Asn Glu Asn Gln Gln
 625 630 635 640
 Pro Ser Glu Ala Ser Lys Glu Glu Lys Glu Ser Asp Asp Phe Ile Asp
 645 650 655
 Ser Leu Pro Asp Tyr Gly Leu Asp Arg Ala Thr Leu Glu Asp His Ile
 660 665 670
 Asn Gln Leu Ala Gln Lys Ala Asn Ile Asp Pro Lys Tyr Leu Ile Phe
 675 680 685
 Gln Pro Glu Gly Val Gln Phe Tyr Asn Lys Asn Gly Glu Leu Val Thr
 690 695 700
 Tyr Asp Ile Lys Thr Leu Gln Gln Ile Asn Pro
 705 710 715

<210> 17
 <211> 77
 <212> PRT
 <213> Streptococcus

<400> 17
 Met His Ser Phe Ser Asn Pro Gly Tyr Pro Tyr Asp Asn Ala Val Thr
 1 5 10 15

Glu Ala Phe Phe Lys Tyr Leu Lys His Arg Gln Ile Asn Arg Lys His
 20 25 30
 Tyr Gln Asn Ile Lys Gln Val Gln Leu Asp Cys Phe Glu Tyr Ile Glu
 35 40 45
 Asn Phe Tyr Asn Asn Tyr Asn Pro His Thr Ala Asn Leu Gly Leu Thr
 50 55 60
 Pro Asn Gln Lys Glu Glu Asn Tyr Phe Asn Ala Ile Lys
 65 70 75

<210> 18
 <211> 86
 <212> PRT
 <213> Streptococcus

<400> 18
 Met Ala Tyr Tyr Gln Ala Cys Thr Glu Lys Asp Ile Ile Arg Ser Met
 1 5 10 15
 Ser Arg Lys Gly Thr Pro Ala Asp Asn Ala Cys Ile Glu Trp Phe His
 20 25 30
 Thr Val Leu Lys Thr Glu Thr Phe Tyr Phe His Asn Arg Arg Lys Tyr
 35 40 45
 Asn Lys Asp Ser Ile Thr Asn Ile Val Lys Asn Tyr Ile Thr Phe Tyr
 50 55 60
 Asn Glu Thr Arg Ile Gln Gln Arg Leu Asn Asp Gln Ser Pro Val Gln
 65 70 75 80
 Tyr Arg Lys Leu Ile Ala
 85

<210> 19
 <211> 126
 <212> PRT
 <213> Streptococcus

<400> 19
 Met Glu Asn His Phe Ile Tyr Gly Tyr Arg Thr Ile Thr Arg Leu Leu
 1 5 10 15
 Lys Lys Ile His Gly Leu Thr Val Asn Thr Lys Lys Val Tyr Arg Ile
 20 25 30
 Met Lys Asn Asn Gly Trp Leu Cys Arg Thr Arg Thr Lys Lys Val Pro
 35 40 45
 Asn Leu Gly Lys Ala Tyr Tyr Leu Thr Asp Asn Lys Leu Ser Arg Asp
 50 55 60
 Phe His Ala Asp Lys Pro Lys Glu Lys Leu Val Thr Asp Ile Thr Tyr
 65 70 75 80
 Leu Tyr Phe Gly Asn Cys Lys Leu Tyr Leu Ser Ser Ile Met Asn Leu
 85 90 95
 Tyr Asn Arg Glu Ile Ile Ala Tyr Thr Ile Ser Asp Cys Gln Asp Thr
 100 105 110
 Asp Phe Val Leu Asp Thr Leu Asn Gln Leu Lys Leu Pro Lys
 115 120 125

<210> 20
 <211> 96
 <212> PRT
 <213> Streptococcus

<400> 20
 Met Val Lys Lys Ala Tyr Ser Trp Glu Thr Lys Leu Ala Cys Ile Asp
 1 5 10 15
 Met Lys Lys Ala Gly Lys Ser Asn Arg Val Ile Met Glu Thr Leu Gly
 20 25 30
 Ile Lys Asn Asn Ser Gln Ile Tyr Thr Trp Met Lys Trp Tyr Glu Asn
 35 40 45
 Glu Glu Leu Tyr Arg Phe His Gln Gly Val Gly Lys Gln Tyr Thr Tyr
 50 55 60
 Gly Lys Gly Leu Glu His Leu Ser Glu Val Glu Gln Leu Gln Leu Gln
 65 70 75 80
 Val Asp Leu Leu Lys Lys Tyr Arg Gly Leu Ile Arg Lys Ser Ile Lys
 85 90 95

<210> 21
 <211> 288
 <212> PRT
 <213> streptococcus

<400> 21
 Ile Arg Tyr Pro Lys Ala Ser Ser Gly Asp Tyr Gly Thr Lys Arg Glu
 1 5 10 15
 Ile Ile Thr Ala Asn Lys Asp Lys Tyr Ser Ile Ser Lys Met Cys Arg
 20 25 30
 Trp Leu Asn Met Pro His Ser Ser Tyr Tyr Tyr Gln Ala Val Glu Ser
 35 40 45
 Val Ser Glu Thr Glu Phe Glu Glu Thr Ile Lys Arg Ile Phe Leu Asp
 50 55 60
 Ser Glu Ser Arg Tyr Gly Ser Arg Lys Ile Lys Ile Cys Leu Asn Asn
 65 70 75 80
 Glu Gly Ile Thr Leu Ser Arg Arg Arg Ile Arg Arg Ile Met Lys Arg
 85 90 95
 Leu Asn Leu Val Ser Val Tyr Gln Lys Ala Thr Phe Lys Pro His Ser
 100 105 110
 Arg Gly Lys Asn Glu Ala Pro Ile Pro Asn His Leu Asp Arg Gln Phe
 115 120 125
 Lys Gln Glu Arg Pro Leu Gln Ala Leu Val Thr Asp Leu Thr Tyr Val
 130 135 140
 Arg Val Gly Asn Arg Trp Ala Tyr Val Cys Leu Ile Ile Asp Leu Tyr
 145 150 155 160
 Asn Arg Glu Ile Ile Gly Leu Ser Leu Gly Trp His Lys Thr Ala Glu
 165 170 175
 Leu Val Lys Gln Ala Ile Gln Ser Ile Pro Tyr Ala Leu Thr Lys Val
 180 185 190
 Lys Met Phe His Ser Asp Arg Gly Lys Glu Phe Asp Asn Gln Leu Ile
 195 200 205
 Asp Glu Ile Leu Glu Ala Phe Gly Ile Thr Arg Ser Leu Ser Gln Ala
 210 215 220
 Gly Tyr Pro Tyr Asp Asn Ala Val Ala Glu Ser Thr Tyr Arg Ala Phe
 225 230 235 240
 Lys Ile Glu Phe Val Tyr Gln Glu Thr Phe Gln Leu Leu Glu Glu Leu
 245 250 255
 Ala Leu Lys Thr Lys Asp Tyr Val His Trp Trp Asn Tyr His Arg Ile
 260 265 270
 His Gly Ser Leu Asn Tyr Gln Thr Pro Met Thr Lys Arg Leu Ile Ala
 275 280 285

```

<210> 22
<211> 5058
<212> DNA
<213> streptococcus

<220>
<221> CDS
<222> (1) ... (663)

<221> CDS
<222> (763) ... (1344)

<221> CDS
<222> (1362) ... (1739)

<221> CDS
<222> (2266) ... (5058)

<400> 22
aat ttg aaa gca gaa tta tct gta gaa gat gag caa tat aca gca aca      48
Asn Leu Lys Ala Glu Leu Ser Val Glu Asp Glu Gln Tyr Thr Ala Thr
  1                5                10                15

gtt tat ggt aaa tct gct cat ggt tca aca cca caa gaa ggt gtt aat      96
Val Tyr Gly Lys Ser Ala His Gly Ser Thr Pro Gln Glu Gly Val Asn
                20                25                30

ggg gcg act tat tta gct ctt tat cta agt caa ttt gat ttt gaa ggt      144
Gly Ala Thr Tyr Leu Ala Leu Tyr Leu Ser Gln Phe Asp Phe Glu Gly
                35                40                45

cct gct cgt gct ttc tta gat gtt aca gcc aac att att cac gaa gac      192
Pro Ala Arg Ala Phe Leu Asp Val Thr Ala Asn Ile Ile His Glu Asp
                50                55                60

ttc tca ggt gaa aaa ctt gga gta gct tat gaa gat gac tgt atg gga      240
Phe Ser Gly Glu Lys Leu Gly Val Ala Tyr Glu Asp Asp Cys Met Gly
  65                70                75                80

cca ttg agc atg aat gca ggt gtc ttc cag ttt gat gaa act aat gat      288
Pro Leu Ser Met Asn Ala Gly Val Phe Gln Phe Asp Glu Thr Asn Asp
                85                90                95

gat aat act atc gct ctt aat ttc cgt tac cca caa ggg aca gat gct      336
Asp Asn Thr Ile Ala Leu Asn Phe Arg Tyr Pro Gln Gly Thr Asp Ala
                100                105                110

aaa act atc caa act aag ctt gag aaa ctt aac gga gtt gaa aaa gtg      384
Lys Thr Ile Gln Thr Lys Leu Glu Lys Leu Asn Gly Val Glu Lys Val
                115                120                125

act ctt tct gac cat gaa cac aca cca cac tat gta cct atg gac gat      432
Thr Leu Ser Asp His Glu His Thr Pro His Tyr Val Pro Met Asp Asp
  130                135                140

```


38 / 63

gaa tta gta tca acc tta cta gct gtc tat gaa aag caa act ggt ctt	480
Glu Leu Val Ser Thr Leu Leu Ala Val Tyr Glu Lys Gln Thr Gly Leu	
145 150 155 160	
aaa gga cat gaa cag gtt att ggt ggt ggg aca ttt ggt cgc tta ctt	528
Lys Gly His Glu Gln Val Ile Gly Gly Gly Thr Phe Gly Arg Leu Leu	
165 170 175	
gaa cgg ggt gtt gca tac ggt gcc atg ttc cca gga gat gaa aac act	576
Glu Arg Gly Val Ala Tyr Gly Ala Met Phe Pro Gly Asp Glu Asn Thr	
180 185 190	
atg cat caa gct aat gag tac atg cct tta gaa aat att ttc cgt tcg	624
Met His Gln Ala Asn Glu Tyr Met Pro Leu Glu Asn Ile Phe Arg Ser	
195 200 205	
gct gct atc tac gca gaa gct atc tat gaa tta atc aaa taaaataatc	673
Ala Ala Ile Tyr Ala Glu Ala Ile Tyr Glu Leu Ile Lys	
210 215 220	
cttaaactaa atatgtgac aatgataaag ggtggggaag acatgaaagt gtctttgctt	733
cttttcataa ggtagattt ggagacttt atg act gac ttg gaa aaa att att	786
Met Thr Asp Leu Glu Lys Ile Ile	
225	
aaa gca ata aaa agt gat tca cag aat caa aat tat aca gaa aat ggt	834
Lys Ala Ile Lys Ser Asp Ser Gln Asn Gln Asn Tyr Thr Glu Asn Gly	
230 235 240 245	
att gat cct ttg ttt gct gct cct aaa aca gct agg atc aat att gtt	882
Ile Asp Pro Leu Phe Ala Ala Pro Lys Thr Ala Arg Ile Asn Ile Val	
250 255 260	
ggc caa gca cct ggt tta aaa act caa gaa gca aga ctc tat tgg aaa	930
Gly Gln Ala Pro Gly Leu Lys Thr Gln Glu Ala Arg Leu Tyr Trp Lys	
265 270 275	
gat aaa tct gga gat cgt cta cgc cag tgg ctt gga gtt gat gaa gag	978
Asp Lys Ser Gly Asp Arg Leu Arg Gln Trp Leu Gly Val Asp Glu Glu	
280 285 290	
aca ttt tac cat tct gga aaa ttt gct gtt tta cct tta gat ttt tat	1026
Thr Phe Tyr His Ser Gly Lys Phe Ala Val Leu Pro Leu Asp Phe Tyr	
295 300 305	
tac cca ggc aaa gga aaa tca gga gat tta ccc cct aga aaa ggt ttt	1074
Tyr Pro Gly Lys Gly Lys Ser Gly Asp Leu Pro Pro Arg Lys Gly Phe	
310 315 320 325	
gcg gag aaa tgg cac cct ctt att tta aaa gaa atg cct aat gtt caa	1122
Ala Glu Lys Trp His Pro Leu Ile Leu Lys Glu Met Pro Asn Val Gln	
330 335 340	
ttg acc ttg cta gtt ggt cag tat gct cag aaa tat tat ctt gga agc	1170
Leu Thr Leu Leu Val Gly Gln Tyr Ala Gln Lys Tyr Tyr Leu Gly Ser	
345 350 355	

tcc gca cat aaa aat cta aca gaa aca gtt aaa gct tac aaa gac tat	1218
Ser Ala His Lys Asn Leu Thr Glu Thr Val Lys Ala Tyr Lys Asp Tyr	
360 365 370	
cta ccc gat tat tta ccc ctg gtt cac cca tca ccg cga aat caa att	1266
Leu Pro Asp Tyr Leu Pro Leu Val His Pro Ser Pro Arg Asn Gln Ile	
375 380 385	
tgg cta aag aag aat cca tgg ttt gaa aaa gat cta atc gtt gat tta	1314
Trp Leu Lys Lys Asn Pro Trp Phe Glu Lys Asp Leu Ile Val Asp Leu	
390 395 400 405	
caa aag ata gta gca gat att tta aaa gat taaggatagg agttggt atg	1364
Gln Lys Ile Val Ala Asp Ile Leu Lys Asp Met	
410 415	
aga gat aat cat cta cac acg tat ttt tcc tat gat tgt caa acg gca	1412
Arg Asp Asn His Leu His Thr Tyr Phe Ser Tyr Asp Cys Gln Thr Ala	
420 425 430	
ttt gag gac tat att aat ggt ttt aca ggt gaa ttt atc acg aca gaa	1460
Phe Glu Asp Tyr Ile Asn Gly Phe Thr Gly Glu Phe Ile Thr Thr Glu	
435 440 445	
cat ttt gat tta tca aat cct tac acc ggt caa gac gat gtt cct gat	1508
His Phe Asp Leu Ser Asn Pro Tyr Thr Gly Gln Asp Asp Val Pro Asp	
450 455 460	
tat agt gct tat tgt caa aaa ata gat tat ctt aat cag aaa tat gga	1556
Tyr Ser Ala Tyr Cys Gln Lys Ile Asp Tyr Leu Asn Gln Lys Tyr Gly	
465 470 475 480	
aat cga ttt aaa aaa gga att gaa atc ggt tat ttt aaa gat agg gaa	1604
Asn Arg Phe Lys Lys Gly Ile Glu Ile Gly Tyr Phe Lys Asp Arg Glu	
485 490 495	
tca gat att tta gat tat tta aaa aat aaa gaa ttt gat tta aaa cta	1652
Ser Asp Ile Leu Asp Tyr Leu Lys Asn Lys Glu Phe Asp Leu Lys Leu	
500 505 510	
ttg tca atc cat cat aat ggt agg tat gat tat ctg caa gaa gaa gct	1700
Leu Ser Ile His His Asn Gly Arg Tyr Asp Tyr Leu Gln Glu Glu Ala	
515 520 525	
ctg aaa gta cca aca aag gga gct ttt agc aga tta ctt taatcgtatg	1749
Leu Lys Val Pro Thr Lys Gly Ala Phe Ser Arg Leu Leu	
530 535 540	
gaatttgcca taggccgtgt ggaagcgcac gttttagctc acttttgatta tggttttcgt	1809
aagttaaact tagatgtaga agatttaaaa ccgtttgaaa cgcaattgaa gcgcattttc	1869
ataaagatgt tatctaagggt gtttagctttt gaactaaata ccaaaccct ttatctatat	1929
gggaatgaaa aacttttatcg ctatgcttta gagatactca aacagcttgg ttgtaaacia	1989
tactctatag gctctgacgg tcatattcct gaacattttt gttatgaatt tgatagactt	2049
caaggtctgc taaaggacta tcaaattgat gaaaatcatt tgatatgagg aaatttttga	2109
taaaaaagct aggcaatatt gcttagcttt tttgtaatgc tattgatagt tttagtgaaa	2169

atttcaaaaa	aataaagaaa	tcatttactt	gttgcaagcg	cttgcgtaaa	ttggtatgat	2229
tttattggta	acaattcatt	aaaaaaggag	aatgat	atg	aaa aga aaa gac tta	2283
				Met	Lys Arg Lys Asp Leu	
					545	
ttt ggt gat	aaa caa act	caa tac acg	att aga aag	tta agt gtt	gga	2331
Phe Gly Asp	Lys Gln Thr	Gln Tyr Thr	Ile Arg Lys	Leu Ser Val	Gly	
	550		555		560	
gta gct tca	ggt aca aca	ggg gta tgt	att ttt ctt	cat agt cca	cag	2379
Val Ala Ser	Val Thr Thr	Gly Val Cys	Ile Phe Leu	His Ser Pro	Gln	
	565		570		575	
gta ttt gct	gaa gaa gta	agt gtt tct	cct gca act	aca gcg att	gca	2427
Val Phe Ala	Glu Glu Val	Ser Val Ser	Pro Ala Thr	Thr Ala Ile	Ala	
580		585			595	
gag tcg aat	att aat cag	ggt gac aac	caa caa tct	act aat tta	aaa	2475
Glu Ser Asn	Ile Asn Gln	Val Asp Asn	Gln Gln Ser	Thr Asn Leu	Lys	
	600		605		610	
gat gac ata	aac tca aac	tct gag acg	ggt gtg aca	ccc tca gat	atg	2523
Asp Asp Ile	Asn Ser Asn	Ser Glu Thr	Val Val Thr	Pro Ser Asp	Met	
	615		620		625	
ccg gat acc	aag caa tta	gta tca gat	gaa act gac	act caa aag	gga	2571
Pro Asp Thr	Lys Gln Leu	Val Ser Asp	Glu Thr Asp	Thr Gln Lys	Gly	
	630		635		640	
gtg aca gag	ccg gat aag	gcg aca agc	ctg ctt gaa	gaa aat aaa	ggt	2619
Val Thr Glu	Pro Asp Lys	Ala Thr Ser	Leu Leu Glu	Glu Glu Asn	Lys Gly	
	645		650		655	
cct gtt tca	gat aaa aat	acc tta gat	tta aaa gta	gca cca tct	aca	2667
Pro Val Ser	Asp Lys Asn	Thr Leu Asp	Leu Lys Val	Ala Pro Ser	Thr	
660		665			675	
ttg caa aat	act ccc gac	aaa act tct	caa gct ata	ggt gct cca	agc	2715
Leu Gln Asn	Thr Pro Asp	Lys Thr Ser	Gln Ala Ile	Gly Ala Pro	Ser	
	680		685		690	
cct acc ttg	aaa gta gct	aat caa gct	cca cgg att	gaa aat ggt	tac	2763
Pro Thr Leu	Lys Val Ala	Asn Gln Ala	Pro Arg Ile	Glu Asn Gly	Tyr	
	695		700		705	
ttt agg cta	cat ctt aaa	gaa ttg cct	caa ggt cat	cct gta gaa	agc	2811
Phe Arg Leu	His Leu Lys	Glu Leu Pro	Gln Gly His	Pro Val Glu	Ser	
	710		715		720	
act gga ctt	tgg ata tgg	gga gat gtt	gat caa ccg	tct agt aat	tgg	2859
Thr Gly Leu	Trp Ile Trp	Gly Asp Val	Asp Gln Pro	Ser Ser Asn	Trp	
	725		730		735	
cca aat ggt	gct atc cct	atg act gat	gct aag aaa	gat gat tac	ggt	2907
Pro Asn Gly	Ala Ile Pro	Met Thr Asp	Ala Lys Lys	Asp Asp Tyr	Gly	
740		745			755	

tat tat gtt gat ttt aaa tta tct gaa aaa caa cga aaa caa ata tct	2955
Tyr Tyr Val Asp Phe Lys Leu Ser Glu Lys Gln Arg Lys Gln Ile Ser	
760 765 770	
ttt tta att aat aac aaa gca ggg aca aat tta agc ggc gat cat cat	3003
Phe Leu Ile Asn Asn Lys Ala Gly Thr Asn Leu Ser Gly Asp His His	
775 780 785	
att cca tta tta cga cct gag atg aac caa gtt tgg att gat gaa aag	3051
Ile Pro Leu Leu Arg Pro Glu Met Asn Gln Val Trp Ile Asp Glu Lys	
790 795 800	
tac ggt ata cat act tat caa ccc ctc aaa gaa ggg tat gtc cgt att	3099
Tyr Gly Ile His Thr Tyr Gln Pro Leu Lys Glu Gly Tyr Val Arg Ile	
805 810 815	
aac tat ttg agt tcc tct agt aac tat gac cac tta tca gca tgg ctc	3147
Asn Tyr Leu Ser Ser Ser Ser Asn Tyr Asp His Leu Ser Ala Trp Leu	
820 825 830 835	
ttt aaa gat gtt gca acc ccy tca aca act tgg cca gat ggt agt aat	3195
Phe Lys Asp Val Ala Thr Xaa Ser Thr Thr Trp Pro Asp Gly Ser Asn	
840 845 850	
ttt gtg aat caa gga cta tat gga agg tat att gat gta tca cta aaa	3243
Phe Val Asn Gln Gly Leu Tyr Gly Arg Tyr Ile Asp Val Ser Leu Lys	
855 860 865	
act aac gcc aaa gag att ggt ttt cta atc tta gat gaa agt aag aca	3291
Thr Asn Ala Lys Glu Ile Gly Phe Leu Ile Leu Asp Glu Ser Lys Thr	
870 875 880	
gga gat gca gtg aaa gtt caa ccc aac gac tat gtt ttt aga gat tta	3339
Gly Asp Ala Val Lys Val Gln Pro Asn Asp Tyr Val Phe Arg Asp Leu	
885 890 895	
gct aac cat aac caa att ttt gta aaa gat aag gat cca aag gtt tat	3387
Ala Asn His Asn Gln Ile Phe Val Lys Asp Lys Asp Pro Lys Val Tyr	
900 905 910 915	
aat aat cct tat tac att gat caa gtg cag cta aag gat gcc caa caa	3435
Asn Asn Pro Tyr Tyr Ile Asp Gln Val Gln Leu Lys Asp Ala Gln Gln	
920 925 930	
att gat tta aca agt att caa gca agt ttt aca act cta gat ggg gta	3483
Ile Asp Leu Thr Ser Ile Gln Ala Ser Phe Thr Thr Leu Asp Gly Val	
935 940 945	
gat aaa act gaa att tta aaa gaa ttg aaa gtg act gat aaa aat caa	3531
Asp Lys Thr Glu Ile Leu Lys Glu Leu Lys Val Thr Asp Lys Asn Gln	
950 955 960	
aat gct ata caa att tct gat atc act ctc gat act agt aaa tct ctt	3579
Asn Ala Ile Gln Ile Ser Asp Ile Thr Leu Asp Thr Ser Lys Ser Leu	
965 970 975	

tta ata atc aaa ggc gac ttt aat cct aaa caa ggt cat ttc aac ata	3627
Leu Ile Ile Lys Gly Asp Phe Asn Pro Lys Gln Gly His Phe Asn Ile	
980 985 990 995	
tct tat aat ggt aac aat gtc atg aca agg caa tct tgg gaa ttt aaa	3675
Ser Tyr Asn Gly Asn Asn Val Met Thr Arg Gln Ser Trp Glu Phe Lys	
1000 1005 1010	
gac caa ctt tat gct tat agt gga aat tta ggt gca gtt ctc aat caa	3723
Asp Gln Leu Tyr Ala Tyr Ser Gly Asn Leu Gly Ala Val Leu Asn Gln	
1015 1020 1025	
gat ggt tca aaa gtt gaa gcc agc ctc tgg tca ccg agt gct gat agt	3771
Asp Gly Ser Lys Val Glu Ala Ser Leu Trp Ser Pro Ser Ala Asp Ser	
1030 1035 1040	
gtc act atg att att tat gac aaa gat aac caa aac agg gtt gta gcg	3819
Val Thr Met Ile Ile Tyr Asp Lys Asp Asn Gln Asn Arg Val Val Ala	
1045 1050 1055	
act acc ccc ctt gtg aaa aat aat aaa ggt gtt tgg cag acg ata ctt	3867
Thr Thr Pro Leu Val Lys Asn Asn Lys Gly Val Trp Gln Thr Ile Leu	
1060 1065 1070 1075	
gat act aaa tta ggt att aaa aac tat act ggt tac tat tat ctt tac	3915
Asp Thr Lys Leu Gly Ile Lys Asn Tyr Thr Gly Tyr Tyr Tyr Leu Tyr	
1080 1085 1090	
gaa ata aaa aga ggt aag gat aag gtt aag att tta gat cct tat gca	3963
Glu Ile Lys Arg Gly Lys Asp Lys Val Lys Ile Leu Asp Pro Tyr Ala	
1095 1100 1105	
aag tca tta gca gag tgg gat agt aat act gtt aat gat gat att aaa	4011
Lys Ser Leu Ala Glu Trp Asp Ser Asn Thr Val Asn Asp Asp Ile Lys	
1110 1115 1120	
acg gct aaa gca gct ttt gta aat cca agt caa ctt gga cct caa aat	4059
Thr Ala Lys Ala Ala Phe Val Asn Pro Ser Gln Leu Gly Pro Gln Asn	
1125 1130 1135	
tta agt ttt gct aaa att gct aat ttt aaa gga aga caa gat gct gtt	4107
Leu Ser Phe Ala Lys Ile Ala Asn Phe Lys Gly Arg Gln Asp Ala Val	
1140 1145 1150 1155	
ata tac gaa gca cat gta aga gac ttc act tct gat cga tct ttg gat	4155
Ile Tyr Glu Ala His Val Arg Asp Phe Thr Ser Asp Arg Ser Leu Asp	
1160 1165 1170	
gga aaa tta aaa aat caa ttt ggt acc ttt gca gcc ttt tca gag aaa	4203
Gly Lys Leu Lys Asn Gln Phe Gly Thr Phe Ala Ala Phe Ser Glu Lys	
1175 1180 1185	
cta gat tat tta cag aaa tta gga gtt aca cac att cag ctt tta ccg	4251
Leu Asp Tyr Leu Gln Lys Leu Gly Val Thr His Ile Gln Leu Leu Pro	
1190 1195 1200	

gta ttg agt tat ttt tat gtt aat gaa atg gat aag tca cgc tca aca 4299
 Val Leu Ser Tyr Phe Tyr Val Asn Glu Met Asp Lys Ser Arg Ser Thr
 1205 1210 1215

gct tac act tcc tca gac aat aat tac aat tgg ggc tat gac cca cag 4347
 Ala Tyr Thr Ser Ser Asp Asn Asn Tyr Asn Trp Gly Tyr Asp Pro Gln
 1220 1225 1230 1235

agc tat ttt gct ctt tct ggg atg tat tca gag aaa cca aaa gat cca 4395
 Ser Tyr Phe Ala Leu Ser Gly Met Tyr Ser Glu Lys Pro Lys Asp Pro
 1240 1245 1250

tca gca cgt atc gcc gaa tta aaa caa tta ata cat gat att cat aaa 4443
 Ser Ala Arg Ile Ala Glu Leu Lys Gln Leu Ile His Asp Ile His Lys
 1255 1260 1265

cgt ggc atg ggg gtt ata ctt gat gtc gtc tat aat cac act gca aaa 4491
 Arg Gly Met Gly Val Ile Leu Asp Val Val Tyr Asn His Thr Ala Lys
 1270 1275 1280

act tat ctc ttt gag gat ata gaa cct aat tat tat cac ttt atg aat 4539
 Thr Tyr Leu Phe Glu Asp Ile Glu Pro Asn Tyr Tyr His Phe Met Asn
 1285 1290 1295

gaa gat ggt tca cca aga gaa agt ttt gga ggg gga cgt tta gga acc 4587
 Glu Asp Gly Ser Pro Arg Glu Ser Phe Gly Gly Gly Arg Leu Gly Thr
 1300 1305 1310 1315

act cat gca atg agt cgt cgt gtt ttg gtt gat tcc att aaa tat ctt 4635
 Thr His Ala Met Ser Arg Arg Val Leu Val Asp Ser Ile Lys Tyr Leu
 1320 1325 1330

aca agt gaa ttt aaa gtt gat ggt ttc cgt ttt gat atg atg gga gat 4683
 Thr Ser Glu Phe Lys Val Asp Gly Phe Arg Phe Asp Met Met Gly Asp
 1335 1340 1345

cat gat gcg gct gcg att gaa tta gct tat aaa gaa gct aaa gct att 4731
 His Asp Ala Ala Ile Glu Leu Ala Tyr Lys Glu Ala Lys Ala Ile
 1350 1355 1360

aat cct aat atg att atg att ggt gag ggc tgg aga aca ttc caa ggc 4779
 Asn Pro Asn Met Ile Met Ile Gly Glu Gly Trp Arg Thr Phe Gln Gly
 1365 1370 1375

gat caa ggt cag ccg gtt aaa cca gct gac caa gat tgg atg aag tca 4827
 Asp Gln Gly Gln Pro Val Lys Pro Ala Asp Gln Asp Trp Met Lys Ser
 1380 1385 1390 1395

acc gat aca gtt ggc gtc ttt tca gat gat att cgt aat agc ttg aaa 4875
 Thr Asp Thr Val Gly Val Phe Ser Asp Asp Ile Arg Asn Ser Leu Lys
 1400 1405 1410

tct ggt ttt cca aat gaa ggt act cca gct ttc atc aca ggt ggc cca 4923
 Ser Gly Phe Pro Asn Glu Gly Thr Pro Ala Phe Ile Thr Gly Gly Pro
 1415 1420 1425

caa tct tta caa ggt att ttt aaa aat atc aaa gca caa cct ggg aat 4971
 Gln Ser Leu Gln Gly Ile Phe Lys Asn Ile Lys Ala Gln Pro Gly Asn
 1430 1435 1440

ttt gaa gca gat tcg cca gga gat gtg gtg cag tat att gct gca cat 5019
 Phe Glu Ala Asp Ser Pro Gly Asp Val Val Gln Tyr Ile Ala Ala His
 1445 1450 1455

gat aac ctt acc ttg cat gat gtg att gca aaa tca att 5058
 Asp Asn Leu Thr Leu His Asp Val Ile Ala Lys Ser Ile
 1460 1465 1470

<210> 23
 <211> 221
 <212> PRT
 <213> streptococcus

<400> 23
 Asn Leu Lys Ala Glu Leu Ser Val Glu Asp Glu Gln Tyr Thr Ala Thr
 1 5 10 15
 Val Tyr Gly Lys Ser Ala His Gly Ser Thr Pro Gln Glu Gly Val Asn
 20 25 30
 Gly Ala Thr Tyr Leu Ala Leu Tyr Leu Ser Gln Phe Asp Phe Glu Gly
 35 40 45
 Pro Ala Arg Ala Phe Leu Asp Val Thr Ala Asn Ile Ile His Glu Asp
 50 55 60
 Phe Ser Gly Glu Lys Leu Gly Val Ala Tyr Glu Asp Asp Cys Met Gly
 65 70 75 80
 Pro Leu Ser Met Asn Ala Gly Val Phe Gln Phe Asp Glu Thr Asn Asp
 85 90 95
 Asp Asn Thr Ile Ala Leu Asn Phe Arg Tyr Pro Gln Gly Thr Asp Ala
 100 105 110
 Lys Thr Ile Gln Thr Lys Leu Glu Lys Leu Asn Gly Val Glu Lys Val
 115 120 125
 Thr Leu Ser Asp His Glu His Thr Pro His Tyr Val Pro Met Asp Asp
 130 135 140
 Glu Leu Val Ser Thr Leu Leu Ala Val Tyr Glu Lys Gln Thr Gly Leu
 145 150 155 160
 Lys Gly His Glu Gln Val Ile Gly Gly Gly Thr Phe Gly Arg Leu Leu
 165 170 175
 Glu Arg Gly Val Ala Tyr Gly Ala Met Phe Pro Gly Asp Glu Asn Thr
 180 185 190
 Met His Gln Ala Asn Glu Tyr Met Pro Leu Glu Asn Ile Phe Arg Ser
 195 200 205
 Ala Ala Ile Tyr Ala Glu Ala Ile Tyr Glu Leu Ile Lys
 210 215 220

<210> 24
 <211> 194
 <212> PRT
 <213> streptococcus

<400> 24

Met Thr Asp Leu Glu Lys Ile Ile Lys Ala Ile Lys Ser Asp Ser Gln
 1 5 10 15
 Asn Gln Asn Tyr Thr Glu Asn Gly Ile Asp Pro Leu Phe Ala Ala Pro
 20 25 30
 Lys Thr Ala Arg Ile Asn Ile Val Gly Gln Ala Pro Gly Leu Lys Thr
 35 40 45
 Gln Glu Ala Arg Leu Tyr Trp Lys Asp Lys Ser Gly Asp Arg Leu Arg
 50 55 60
 Gln Trp Leu Gly Val Asp Glu Glu Thr Phe Tyr His Ser Gly Lys Phe
 65 70 75 80
 Ala Val Leu Pro Leu Asp Phe Tyr Tyr Pro Gly Lys Gly Lys Ser Gly
 85 90 95
 Asp Leu Pro Pro Arg Lys Gly Phe Ala Glu Lys Trp His Pro Leu Ile
 100 105 110
 Leu Lys Glu Met Pro Asn Val Gln Leu Thr Leu Leu Val Gly Gln Tyr
 115 120 125
 Ala Gln Lys Tyr Tyr Leu Gly Ser Ser Ala His Lys Asn Leu Thr Glu
 130 135 140
 Thr Val Lys Ala Tyr Lys Asp Tyr Leu Pro Asp Tyr Leu Pro Leu Val
 145 150 155 160
 His Pro Ser Pro Arg Asn Gln Ile Trp Leu Lys Lys Asn Pro Trp Phe
 165 170 175
 Glu Lys Asp Leu Ile Val Asp Leu Gln Lys Ile Val Ala Asp Ile Leu
 180 185 190
 Lys Asp

<210> 25
 <211> 126
 <212> PRT
 <213> streptococcus

<400> 25
 Met Arg Asp Asn His Leu His Thr Tyr Phe Ser Tyr Asp Cys Gln Thr
 1 5 10 15
 Ala Phe Glu Asp Tyr Ile Asn Gly Phe Thr Gly Glu Phe Ile Thr Thr
 20 25 30
 Glu His Phe Asp Leu Ser Asn Pro Tyr Thr Gly Gln Asp Asp Val Pro
 35 40 45
 Asp Tyr Ser Ala Tyr Cys Gln Lys Ile Asp Tyr Leu Asn Gln Lys Tyr
 50 55 60
 Gly Asn Arg Phe Lys Lys Gly Ile Glu Ile Gly Tyr Phe Lys Asp Arg
 65 70 75 80
 Glu Ser Asp Ile Leu Asp Tyr Leu Lys Asn Lys Glu Phe Asp Leu Lys
 85 90 95
 Leu Leu Ser Ile His His Asn Gly Arg Tyr Asp Tyr Leu Gln Glu Glu
 100 105 110
 Ala Leu Lys Val Pro Thr Lys Gly Ala Phe Ser Arg Leu Leu
 115 120 125

<210> 26
 <211> 931
 <212> PRT
 <213> streptococcus

<400> 26

46 / 63

Met Lys Arg Lys Asp Leu Phe Gly Asp Lys Gln Thr Gln Tyr Thr Ile
 1 5 10 15
 Arg Lys Leu Ser Val Gly Val Ala Ser Val Thr Thr Gly Val Cys Ile
 20 25 30
 Phe Leu His Ser Pro Gln Val Phe Ala Glu Glu Val Ser Val Ser Pro
 35 40 45
 Ala Thr Thr Ala Ile Ala Glu Ser Asn Ile Asn Gln Val Asp Asn Gln
 50 55 60
 Gln Ser Thr Asn Leu Lys Asp Asp Ile Asn Ser Asn Ser Glu Thr Val
 65 70 75 80
 Val Thr Pro Ser Asp Met Pro Asp Thr Lys Gln Leu Val Ser Asp Glu
 85 90 95
 Thr Asp Thr Gln Lys Gly Val Thr Glu Pro Asp Lys Ala Thr Ser Leu
 100 105 110
 Leu Glu Glu Asn Lys Gly Pro Val Ser Asp Lys Asn Thr Leu Asp Leu
 115 120 125
 Lys Val Ala Pro Ser Thr Leu Gln Asn Thr Pro Asp Lys Thr Ser Gln
 130 135 140
 Ala Ile Gly Ala Pro Ser Pro Thr Leu Lys Val Ala Asn Gln Ala Pro
 145 150 155 160
 Arg Ile Glu Asn Gly Tyr Phe Arg Leu His Leu Lys Glu Leu Pro Gln
 165 170 175
 Gly His Pro Val Glu Ser Thr Gly Leu Trp Ile Trp Gly Asp Val Asp
 180 185 190
 Gln Pro Ser Ser Asn Trp Pro Asn Gly Ala Ile Pro Met Thr Asp Ala
 195 200 205
 Lys Lys Asp Asp Tyr Gly Tyr Tyr Val Asp Phe Lys Leu Ser Glu Lys
 210 215 220
 Gln Arg Lys Gln Ile Ser Phe Leu Ile Asn Asn Lys Ala Gly Thr Asn
 225 230 235 240
 Leu Ser Gly Asp His His Ile Pro Leu Leu Arg Pro Glu Met Asn Gln
 245 250 255
 Val Trp Ile Asp Glu Lys Tyr Gly Ile His Thr Tyr Gln Pro Leu Lys
 260 265 270
 Glu Gly Tyr Val Arg Ile Asn Tyr Leu Ser Ser Ser Ser Asn Tyr Asp
 275 280 285
 His Leu Ser Ala Trp Leu Phe Lys Asp Val Ala Thr Xaa Ser Thr Thr
 290 295 300
 Trp Pro Asp Gly Ser Asn Phe Val Asn Gln Gly Leu Tyr Gly Arg Tyr
 305 310 315 320
 Ile Asp Val Ser Leu Lys Thr Asn Ala Lys Glu Ile Gly Phe Leu Ile
 325 330 335
 Leu Asp Glu Ser Lys Thr Gly Asp Ala Val Lys Val Gln Pro Asn Asp
 340 345 350
 Tyr Val Phe Arg Asp Leu Ala Asn His Asn Gln Ile Phe Val Lys Asp
 355 360 365
 Lys Asp Pro Lys Val Tyr Asn Asn Pro Tyr Tyr Ile Asp Gln Val Gln
 370 375 380
 Leu Lys Asp Ala Gln Gln Ile Asp Leu Thr Ser Ile Gln Ala Ser Phe
 385 390 395 400
 Thr Thr Leu Asp Gly Val Asp Lys Thr Glu Ile Leu Lys Glu Leu Lys
 405 410 415
 Val Thr Asp Lys Asn Gln Asn Ala Ile Gln Ile Ser Asp Ile Thr Leu
 420 425 430
 Asp Thr Ser Lys Ser Leu Leu Ile Ile Lys Gly Asp Phe Asn Pro Lys
 435 440 445

47 / 63

Gln Gly His Phe Asn Ile Ser Tyr Asn Gly Asn Asn Val Met Thr Arg
 450 455 460
 Gln Ser Trp Glu Phe Lys Asp Gln Leu Tyr Ala Tyr Ser Gly Asn Leu
 465 470 475 480
 Gly Ala Val Leu Asn Gln Asp Gly Ser Lys Val Glu Ala Ser Leu Trp
 485 490 495
 Ser Pro Ser Ala Asp Ser Val Thr Met Ile Ile Tyr Asp Lys Asp Asn
 500 505 510
 Gln Asn Arg Val Val Ala Thr Thr Pro Leu Val Lys Asn Asn Lys Gly
 515 520 525
 Val Trp Gln Thr Ile Leu Asp Thr Lys Leu Gly Ile Lys Asn Tyr Thr
 530 535 540
 Gly Tyr Tyr Tyr Leu Tyr Glu Ile Lys Arg Gly Lys Asp Lys Val Lys
 545 550 555 560
 Ile Leu Asp Pro Tyr Ala Lys Ser Leu Ala Glu Trp Asp Ser Asn Thr
 565 570 575
 Val Asn Asp Asp Ile Lys Thr Ala Lys Ala Ala Phe Val Asn Pro Ser
 580 585 590
 Gln Leu Gly Pro Gln Asn Leu Ser Phe Ala Lys Ile Ala Asn Phe Lys
 595 600 605
 Gly Arg Gln Asp Ala Val Ile Tyr Glu Ala His Val Arg Asp Phe Thr
 610 615 620
 Ser Asp Arg Ser Leu Asp Gly Lys Leu Lys Asn Gln Phe Gly Thr Phe
 625 630 635 640
 Ala Ala Phe Ser Glu Lys Leu Asp Tyr Leu Gln Lys Leu Gly Val Thr
 645 650 655
 His Ile Gln Leu Leu Pro Val Leu Ser Tyr Phe Tyr Val Asn Glu Met
 660 665 670
 Asp Lys Ser Arg Ser Thr Ala Tyr Thr Ser Ser Asp Asn Asn Tyr Asn
 675 680 685
 Trp Gly Tyr Asp Pro Gln Ser Tyr Phe Ala Leu Ser Gly Met Tyr Ser
 690 695 700
 Glu Lys Pro Lys Asp Pro Ser Ala Arg Ile Ala Glu Leu Lys Gln Leu
 705 710 715 720
 Ile His Asp Ile His Lys Arg Gly Met Gly Val Ile Leu Asp Val Val
 725 730 735
 Tyr Asn His Thr Ala Lys Thr Tyr Leu Phe Glu Asp Ile Glu Pro Asn
 740 745 750
 Tyr Tyr His Phe Met Asn Glu Asp Gly Ser Pro Arg Glu Ser Phe Gly
 755 760 765
 Gly Gly Arg Leu Gly Thr Thr His Ala Met Ser Arg Arg Val Leu Val
 770 775 780
 Asp Ser Ile Lys Tyr Leu Thr Ser Glu Phe Lys Val Asp Gly Phe Arg
 785 790 795 800
 Phe Asp Met Met Gly Asp His Asp Ala Ala Ile Glu Leu Ala Tyr
 805 810 815
 Lys Glu Ala Lys Ala Ile Asn Pro Asn Met Ile Met Ile Gly Glu Gly
 820 825 830
 Trp Arg Thr Phe Gln Gly Asp Gln Gly Gln Pro Val Lys Pro Ala Asp
 835 840 845
 Gln Asp Trp Met Lys Ser Thr Asp Thr Val Gly Val Phe Ser Asp Asp
 850 855 860
 Ile Arg Asn Ser Leu Lys Ser Gly Phe Pro Asn Glu Gly Thr Pro Ala
 865 870 875 880
 Phe Ile Thr Gly Gly Pro Gln Ser Leu Gln Gly Ile Phe Lys Asn Ile
 885 890 895

Lys Ala Gln Pro Gly Asn Phe Glu Ala Asp Ser Pro Gly Asp Val Val
 900 905 910
 Gln Tyr Ile Ala Ala His Asp Asn Leu Thr Leu His Asp Val Ile Ala
 915 920 925
 Lys Ser Ile
 930

<210> 27
 <211> 5607
 <212> DNA
 <213> streptococcus

 <220>
 <221> CDS
 <222> (2)...(301)

<400> 27
 a att caa agt ttg aca gaa ggt caa ctt cgt tct gat atc cct gag ttc 49
 Ile Gln Ser Leu Thr Glu Gly Gln Leu Arg Ser Asp Ile Pro Glu Phe
 1 5 10 15

 cgt gct ggt gat act gta cgt gtt cac gct aaa gtt gtt gaa ggt act 97
 Arg Ala Gly Asp Thr Val Arg Val His Ala Lys Val Val Glu Gly Thr
 20 25 30

 cgc gaa cgt att cag atc ttt gaa ggt gtt gtt atc tca cgt aaa ggt 145
 Arg Glu Arg Ile Gln Ile Phe Glu Gly Val Val Ile Ser Arg Lys Gly
 35 40 45

 caa gga atc tca gaa atg tac aca gta cgt aaa att tct ggt ggt atc 193
 Gln Gly Ile Ser Glu Met Tyr Thr Val Arg Lys Ile Ser Gly Gly Ile
 50 55 60

 ggt gta gag cgt aca ttc cca att cac act cct cgt gtt gat aaa atc 241
 Gly Val Glu Arg Thr Phe Pro Ile His Thr Pro Arg Val Asp Lys Ile
 65 70 75 80

 gaa gtt gtt cgt tat ggt aaa gta cgt cgt gct aaa ctt tac tac tta 289
 Glu Val Val Arg Tyr Gly Lys Val Arg Arg Ala Lys Leu Tyr Tyr Leu
 85 90 95

 cgc gca ttg caa ggtaaagctg cacgtattaa agaaatccgt cgtaaatttt 341
 Arg Ala Leu Gln
 100

gatgatcaga ttttaaaaaat gcttggttgt ttgaggatag taactatggt ttaaaaactgg 401
 acaaccaaga cgtaaaaaat ctgcctgtgg gcagtttttt tactaggtcc ccttagttca 461
 atggatataa caactccctc ctaaggagta attgctgggt cgattccggc aggggacata 521
 ttcattgcat gtaaatagcg gtttagagct attttgcccc aaatttctct gattaagtgt 581
 atcgttcccta tctttttggt cttgtaattg atgtgcgtaa acttctaaag tgatatttaa 641
 attctcgtga tctaaaactt gagagatgga aattagatag cttgcaaag tatgcctgag 701
 agagtgcact cgtacctcgc gaccagttat ttttcggata gttttattga ctgcattatt 761
 tgaagtttg tcaataatc tgtcgttttt attttttgta aattcatgca aaaaaataa 821
 tgtatcattg tcaattgga tatttctgat actacttttg ttttttggtg gcaggtatct 881
 ttggttgaaa tgataatccc aagttttatt aattgataaa tatttgttag tgtaatcaat 941
 atcattaact gttaaacta aacattcagc gaagcgcagc ccagtttttag cgatgaggta 1001

taacgctgca	tacgattgat	gttgtgattt	ttctttacaa	atTTTTatca	agcgtaagta	1061
ttcattgggt	tcaagaaatt	ttatctctat	ttacgcccct	tatTTTTtgc	tttaacctta	1121
gtgaataaac	aaaaattttt	ttctatata	ccctcgtgaa	cagccatgga	tacgcaggct	1181
ttacatgta	tgttaaaacg	ctttactgta	tcttgccat	gcgtttgact	ataatgattt	1241
atgacttggt	gatatttagt	ggaagtaata	ttgcaaagta	atatatttcc	tatttatatgt	1301
ttatacgata	ttcgatattc	ccaccggtt	tcgctgttac	ggaaatacgc	cattgatata	1361
ctccacatta	gctaaagAAC	aggggtgttca	aggctacctt	gatggaaaag	gctctcttag	1421
agatatttgt	aatgggtatg	atatctcaag	tcgctctggt	ctccaaaagt	ggataaaaacg	1481
gtatactagt	ggtgaagact	tgaaagccac	tagtagagga	tatagccgta	tgaaacaagg	1541
aaggcaagcc	acatttgaag	aacgtgtaga	gattgttaac	tacaccattg	cccattggaa	1601
agactatcaa	gcagctattg	agaagtttgg	tgtttcctac	caacaaattt	attcttgggt	1661
gcgtaagcct	gagaagaatg	gctcacaagg	tttggttgat	agacgtgtga	aagggttgga	1721
gagtaggctt	gatttaaccg	agattgagca	actttaactc	aagattaaac	aattggagga	1781
acgtaatcgt	ctcttagaaa	tcgaggttag	tttactaaaa	aagttagaag	acatcaaacg	1841
aggaacaga	cggtaaagact	aggtaagcat	ttagcggagt	tccaagtaat	caagaattat	1901
tacgatgagg	aatctaattg	gcctattcag	gccttatgcc	aactcttgaa	ggggtctcgt	1961
tcaggctatt	acaagtggct	caatcgtcaa	aaaacagatt	ttgagacaaa	aaatacaagg	2021
ctaattggcta	aaatcaagga	acttcgtaga	ctctacaatg	gtatcttagg	ttatcgccgt	2081
atgacaacat	ttattaatcg	tcaacttggg	acaacttaaa	acaagaaacg	gattcgttga	2141
ttgatgaaca	ttctggggat	tagttcagtc	attcgtcgtg	ttagccatgc	ttgtacaaaa	2201
gctggtgaca	gattttacga	agaaaatatt	cttaatcgtg	aatttacagc	cacagctcat	2261
aaccagaaat	ggtgcacaga	tgtcacctat	cttcaatac	gtctgggagc	taaagcttat	2321
ctcagtgcga	ttaaagacct	gtataacggg	tctattatcg	cttatgagat	tagtcacaac	2381
aatgaaatcc	acttgttatg	aagaccatta	aaaaggggct	agagctcaat	ccaggagcca	2441
cacctatcat	ccatagcgat	tgaggtagtc	aatatacttc	caaagaatac	cgttatatca	2501
tacaacaagc	tggtctgacc	ttatccatgt	cccggattgg	caaatgtatt	gataatgcac	2561
caactgaaag	tttctttggg	tttttcaaga	ctgagtotta	ccaccttaag	aaatacaact	2621
cttatgatga	gttgggtcaat	gatgtggcac	gttatatoga	attctacaac	acacaacggt	2681
atcaatcaaa	attaaacaac	ctgactcctc	tagaattcag	gaatcagggt	gcataactta	2741
tcttttatta	tttgactgtc	tacttgacag	ggagccggtc	agattgctta	acctttctaa	2801
atTTgctaaa	atagctacaa	gaaaacgagc	catttaatgc	ttattttctta	tactgtcttg	2861
cctcacgcca	tcctcgacca	aaaattgagc	gattggcttt	ttgtttcatt	aaacgatgat	2921
atTTccatat	tcatcagttt	gttttccgag	agccatcaaa	gcttcgataa	ggctgataat	2981
tccaggaata	aaggtaatat	taaaaataat	atataaaaaa	acctggccta	tttttctctg	3041
gtaaaattta	tgcgctccaa	tgccgcccaa	aagaacgtta	ataaaacata	aactactatg	3101
ttagcataag	actttatttt	tacaactgaa	tttcatataa	atggattaga	gtaagggata	3161
aaagaaatta	gcatagctct	tttgaaaata	aaaaaattaa	tataatatgg	aaaaattttt	3221
atTTcataaa	cgtttcataa	aaggatgtta	atctagtatt	taggcaacac	tattttgtca	3281
ctgggtgcta	gtaacttata	gattgataat	tttactagta	aacgtaattc	ttcgctttaa	3341
gagttaaatg	tctatttatt	gtaagctaaa	ttgggaggtg	aacttatgta	aaattagata	3401
ggtactgtca	agtacgggat	gattattgaa	acagccagta	tgcatcataa	aatctgtatt	3461
gcttaataac	tatttcctta	accagacatc	agttcattgt	ttatcatcgc	taccctaagt	3521
ctagtttttt	caatagagca	ttaggtagtt	tttgataata	aaactatata	aacatgagaa	3581
ttagatttctg	tattgcattc	ttcataatga	gttatTTgag	atTTtccttt	gaataaatag	3641
atacgaaatt	cagtaacttc	atatataaac	ggctctatca	ttgagatagt	ttgtcaaattg	3701
aagaaatttt	taatggaaat	agttttaaa	acattagttg	tagggcatgt	aaaaatatta	3761
atccagtgga	tgcaatagtt	goggagtaaa	aatagagagg	agtaattagg	aagtgataaa	3821
aaatgctata	gcatatatta	ccagaaaaaa	aaatagaaca	cttattatat	ttgctatttt	3881
aacaattggt	ctttcttgct	tgtattcatg	tttaacaata	atgaaatcaa	gtaatgaaat	3941
agaaaaggct	ttatatgaaa	gttctaattc	ttcaatatca	attacaaaaa	aagatggtaa	4001
atatttttaat	attaatcaat	ttaagaatat	tgaaaaaata	aaagaggttg	aagaaaaaat	4061
atTTcaatat	gatggattag	caaaattgaa	agatcttaaa	gtagtttagtg	gtgagcaaa	4121
tataaataga	gaagatttat	ctgacgaatt	taaaaatggt	gtttcactag	aagctacaag	4181
taactactaaa	agaaatcttt	tatttagtagt	tgaggtattt	agttttaaag	aaggaaaaaa	4241
tatagaagaa	aatgataaga	attcaattct	tgttcatgaa	gaatttgcta	aacaaaaaca	4301
actaaaattg	ggtgatgaaa	ttgatcttga	attactagat	acggaaaaaa	gtggaaaaat	4361

```

aaaaagtcac aaatttaaaa ttataggaat cttttctggt aaaaaacagg aaacatatac 4421
aggattatca tctgatttta gcgaaaatat ggtttttgta gattattcaa cttagccaaga 4481
aatattaat aaatcagaga ataatagaat tgcaaataaa attttaatgt attctggtag 4541
tttagaatct acagagcttg ccttaaacaa attgaaagac tttaaaattg ataagtcaaa 4601
gtattctatt aagaaagata ataaagcatt cgaagagtct ttagagtcag tgagtggaaat 4661
aaaacatata attaaaataa tgacttattc gattatgtta ggtggaatag ttgttctttc 4721
attaatcttg attctatggg taagagaaaag aatttatgaa ataggtatat ttttatctat 4781
tggaacaact aagatacaaa ttataaggca atttatattt gagttaatat tcatatcaat 4841
accaagtata atactctcct tatttttagg gaatctacta ttaaaagtaa ttgtagaagg 4901
atttattaac tcagaaaact caatgatttt cgggtggaagt ttaataaata aaagcagttt 4961
tatgtttaaac ataacaacac ttgcagaaaag ttatttaata ttaataagta ttattgtttt 5021
atcagttgta atggcctctt cattaatatt atttaagaaa ccacaagaaa tattatcaaa 5081
aataagttag gagcaaataa tggatatatt agaaataaag aatgtaaatt acagttacgc 5141
aaattctaaa gaaaaagttt tgtcaggagt aaatcaaaaa tttgaaacttg gaaagtttta 5201
tgcgatagta gggaaagtcag gaacaggaaa atccacactt ctttcccttac ttgcaggact 5261
tgataaagtt caaacaggaa aaatccttgtt taagaatgaa gatatagaaa agaaaggata 5321
tagtaatcac agaaaaaata atatatcttt cgtatttcaa aattataatt taatagatta 5381
tttatcgccg attgaaaata ttagactagt aaataaatca gtagatgaga gtatcttgtt 5441
cgaattaggt ttagataaaa aacaaataaa aagaaatggt atgaaattat ctggtgtgtca 5501
gcaacaaagg gtagctattg ctagggcact ggtatcagat gccccataaa tactagctga 5561
tgagcctacc ggtaacctag acagtgttac tgctggagaa ataatt 5607

```

<210> 28

```

<211> 111
<212> PRT
<213> streptococcus

```

<400> 28

```

Ile Gln Ser Leu Thr Glu Gly Gln Leu Arg Ser Asp Ile Pro Glu Phe
 1          5          10          15
Arg Ala Gly Asp Thr Val Arg Val His Ala Lys Val Val Glu Gly Thr
 20          25          30
Arg Glu Arg Ile Gln Ile Phe Glu Gly Val Val Ile Ser Arg Lys Gly
 35          40          45
Gln Gly Ile Ser Glu Met Tyr Thr Val Arg Lys Ile Ser Gly Gly Ile
 50          55          60
Gly Val Glu Arg Thr Phe Pro Ile His Thr Pro Arg Val Asp Lys Ile
 65          70          75          80
Glu Val Val Arg Tyr Gly Lys Val Arg Arg Ala Lys Leu Tyr Tyr Leu
 85          90          95
Arg Ala Leu Gln Gly Lys Ala Ala Arg Ile Lys Glu Ile Arg Arg
 100          105          110

```

```

<210> 29
<211> 173
<212> PRT
<213> streptococcus

```

<400> 29

```

Met Arg Phe Ala Glu Cys Leu Gly Leu Thr Val Asn Asp Ile Asp Tyr
 1          5          10          15
Thr Asn Lys Tyr Leu Ser Ile Asn Lys Thr Trp Asp Tyr His Phe Asn
 20          25          30
Gln Arg Tyr Leu Pro Thr Lys Asn Lys Ser Ser Ile Arg Asn Ile Pro
 35          40          45
Ile Asp Asn Asp Thr Leu Phe Phe Leu His Glu Phe Thr Lys Asn Lys

```

50					55					60					
Asn	Asp	Arg	Leu	Phe	Asp	Lys	Leu	Ser	Asn	Asn	Ala	Val	Asn	Lys	Thr
65					70					75					80
Ile	Arg	Lys	Ile	Thr	Gly	Arg	Glu	Val	Arg	Val	His	Ser	Leu	Arg	His
				85					90					95	
Thr	Phe	Ala	Ser	Tyr	Leu	Ile	Ser	Ile	Ser	Gln	Val	Leu	Asp	His	Glu
			100					105					110		
Asn	Leu	Asn	Ile	Thr	Leu	Glu	Val	Tyr	Ala	His	Gln	Leu	Gln	Glu	Gln
		115						120					125		
Lys	Asp	Arg	Asn	Asp	Lys	Leu	Asn	Gln	Arg	Asn	Leu	Gly	Gln	Asn	Ser
	130					135						140			
Ser	Lys	Pro	Leu	Phe	Thr	Cys	Asn	Glu	Tyr	Val	Pro	Cys	Arg	Asn	Arg
145					150					155					160
Thr	Ser	Asn	Tyr	Ser	Leu	Gly	Gly	Ser	Cys	Tyr	Ile	His			
				165					170						

<210> 30
 <211> 389
 <212> PRT
 <213> streptococcus

<400>	30														
Met	Lys	Ser	Ser	Asn	Glu	Ile	Glu	Lys	Ala	Leu	Tyr	Glu	Ser	Ser	Asn
1				5					10					15	
Ser	Ser	Ile	Ser	Ile	Thr	Lys	Lys	Asp	Gly	Lys	Tyr	Phe	Asn	Ile	Asn
			20					25					30		
Gln	Phe	Lys	Asn	Ile	Glu	Lys	Ile	Lys	Glu	Val	Glu	Glu	Lys	Ile	Phe
		35					40					45			
Gln	Tyr	Asp	Gly	Leu	Ala	Lys	Leu	Lys	Asp	Leu	Lys	Val	Val	Ser	Gly
	50					55					60				
Glu	Gln	Ser	Ile	Asn	Arg	Glu	Asp	Leu	Ser	Asp	Glu	Phe	Lys	Asn	Val
65				70						75					80
Val	Ser	Leu	Glu	Ala	Thr	Ser	Asn	Thr	Lys	Arg	Asn	Leu	Leu	Phe	Ser
				85				90						95	
Ser	Gly	Val	Phe	Ser	Phe	Lys	Glu	Gly	Lys	Asn	Ile	Glu	Glu	Asn	Asp
			100					105						110	
Lys	Asn	Ser	Ile	Leu	Val	His	Glu	Glu	Phe	Ala	Lys	Gln	Asn	Lys	Leu
		115						120						125	
Lys	Leu	Gly	Asp	Glu	Ile	Asp	Leu	Glu	Leu	Leu	Asp	Thr	Glu	Lys	Ser
		130				135						140			
Gly	Lys	Ile	Lys	Ser	His	Lys	Phe	Lys	Ile	Ile	Gly	Ile	Phe	Ser	Gly
145					150						155				160
Lys	Lys	Gln	Glu	Thr	Tyr	Thr	Gly	Leu	Ser	Ser	Asp	Phe	Ser	Glu	Asn
				165					170					175	
Met	Val	Phe	Val	Asp	Tyr	Ser	Thr	Ser	Gln	Glu	Ile	Leu	Asn	Lys	Ser
			180						185					190	
Glu	Asn	Asn	Arg	Ile	Ala	Asn	Lys	Ile	Leu	Met	Tyr	Ser	Gly	Ser	Leu
		195					200						205		
Glu	Ser	Thr	Glu	Leu	Ala	Leu	Asn	Lys	Leu	Lys	Asp	Phe	Lys	Ile	Asp
		210					215					220			
Lys	Ser	Lys	Tyr	Ser	Ile	Lys	Lys	Asp	Asn	Lys	Ala	Phe	Glu	Glu	Ser
225					230					235					240
Leu	Glu	Ser	Val	Ser	Gly	Ile	Lys	His	Ile	Ile	Lys	Ile	Met	Thr	Tyr
				245						250				255	
Ser	Ile	Met	Leu	Gly	Gly	Ile	Val	Val	Leu	Ser	Leu	Ile	Leu	Ile	Leu
			260					265						270	

Trp Leu Arg Glu Arg Ile Tyr Glu Ile Gly Ile Phe Leu Ser Ile Gly
 275 280 285
 Thr Thr Lys Ile Gln Ile Ile Arg Gln Phe Ile Phe Glu Leu Ile Phe
 290 295 300
 Ile Ser Ile Pro Ser Ile Ile Ser Ser Leu Phe Leu Gly Asn Leu Leu
 305 310 315 320
 Leu Lys Val Ile Val Glu Gly Phe Ile Asn Ser Glu Asn Ser Met Ile
 325 330 335
 Phe Gly Gly Ser Leu Ile Asn Lys Ser Ser Phe Met Leu Asn Ile Thr
 340 345 350
 Thr Leu Ala Glu Ser Tyr Leu Ile Leu Ile Ser Ile Ile Val Leu Ser
 355 360 365
 Val Val Met Ala Ser Ser Leu Ile Leu Phe Lys Lys Pro Gln Glu Ile
 370 375 380
 Leu Ser Lys Ile Ser
 385

<210> 31
 <211> 169
 <212> PRT
 <213> streptococcus

<400> 31
 Met Asp Ile Leu Glu Ile Lys Asn Val Asn Tyr Ser Tyr Ala Asn Ser
 1 5 10 15
 Lys Glu Lys Val Leu Ser Gly Val Asn Gln Lys Phe Glu Leu Gly Lys
 20 25 30
 Phe Tyr Ala Ile Val Gly Lys Ser Gly Thr Gly Lys Ser Thr Leu Leu
 35 40 45
 Ser Leu Leu Ala Gly Leu Asp Lys Val Gln Thr Gly Lys Ile Leu Phe
 50 55 60
 Lys Asn Glu Asp Ile Glu Lys Lys Gly Tyr Ser Asn His Arg Lys Asn
 65 70 75 80
 Asn Ile Ser Leu Val Phe Gln Asn Tyr Asn Leu Ile Asp Tyr Leu Ser
 85 90 95
 Pro Ile Glu Asn Ile Arg Leu Val Asn Lys Ser Val Asp Glu Ser Ile
 100 105 110
 Leu Phe Glu Leu Gly Leu Asp Lys Lys Gln Ile Lys Arg Asn Val Met
 115 120 125
 Lys Leu Ser Gly Gly Gln Gln Gln Arg Val Ala Ile Ala Arg Ala Leu
 130 135 140
 Val Ser Asp Ala Pro Ile Ile Leu Ala Asp Glu Pro Thr Gly Asn Leu
 145 150 155 160
 Asp Ser Val Thr Ala Gly Glu Ile Ile
 165

<210> 32
 <211> 4171
 <212> DNA
 <213> Streptococcus

<400> 32
 catatgacaa tatttttcaa agtctacatc acttactcgc ctgtcgtgga aaatctggca 60
 atacattaat cgaccaatta gttgctgatg gtttacttca tgcagataat cactaccatt 120
 ttttcaatgg gaagtctctg gccactttca atactaacca attgattcgc gaagttgtct 180
 atgttgaaat atccttagat actatgtcta gtgggtgaaca tgatttagta aaagttaaca 240

ttatcagacc	cactaccgag	catactatcc	ccacgatgat	gacagctagc	ccctatcatc	300
aaggatcaaa	tgatcctgcc	gcagaccaaa	aaacataacca	aatggagggg	gcgctagcag	360
ttaaacagcc	taaacacata	caagttgaca	caaaaccatt	taaagaagaa	gtaaaacatc	420
cttcaaaatt	acccatcagc	cctgcaactg	aaagcttcac	acacattgac	agttatagtc	480
tcaatgacta	ttttctttct	cgtggttttg	ctaatatata	cgtttcaggt	gtgggtactg	540
ctggctctac	gggtttcatg	accagtgggg	attaccaaca	aatacaaagc	tttaaagcag	600
tcattgattg	gttaaattgg	aaggttactg	cattcacaag	tcataaacga	gataaacaag	660
tcaaggctga	ttgggtcaaac	ggccttgtag	caaccacagg	taaattcttat	ctcggtagca	720
tgtcaactgg	tttagcaaca	actggcggtg	aggggctgaa	agtcattatc	gctgaagccg	780
caatctccac	atgggtatgat	tattatcgtag	aaaatgggct	tgtgtgtagt	ccagggcgtc	840
accccggtga	agatttagac	gttttaacag	aattaacata	ctcacgaaac	ctcttagctg	900
gtgattacat	caaaaacaac	gattgctatc	aagcattggt	aatgaacaa	tcaaaagcaa	960
ttgaccgtca	aagtggggat	tacaaccaat	actggcatga	ccgtaattac	ctaactcacg	1020
tcaataatgt	caaaagtcca	gtagtttaca	ctcatggact	acaggattgg	aatgttaagc	1080
caagacatgt	ctacaaagt	ttcaatgcat	tgctcaaac	catcaaaaaa	cacctttttt	1140
tacatcaagg	tcaacatggt	tatatgcata	attggcagtc	gattgatttt	cgtgaaagca	1200
tgaatgcctt	actaagccaa	gaactacttg	gcattgacaa	tcatttccaa	ttagaagagg	1260
tcatttgcca	agataaact	actgagcaaa	cttggcaagt	tttagatgct	ttcggaggaa	1320
accatcaaga	gcaaattgg	ttaggtgata	gtaaaaaact	tattgataac	cattatgaca	1380
aagaagcctt	tgatacttat	tgtaaaagact	tcaatgtggt	caaaaatgat	cttttcaagg	1440
gaaataataa	aaccaatcaa	atcactatta	atcttctct	aaagaaaaat	tatctcttga	1500
atggacagt	caaactccat	ctacgtggt	aaactagtga	caaaaaggcc	attttatcag	1560
cccaaatctt	agacatgggt	cctaaaaaac	gattcaaaga	tacaccaacc	atcaaatctt	1620
taaacagcct	tgataatggt	aaaaattttg	ccagagaagc	tttacgtgaa	ctcccgttta	1680
ctaaagatca	ttatcgtgtc	atcagtaaag	gtgtcttgaa	ccttcaaaaat	cgtacagact	1740
tacttacaat	tgaggctatc	gagccagaac	aatggtttga	tatcgagttt	agcctccaac	1800
caagtatata	tcaattgagt	aaaggtgata	atctaaggat	tatcctttat	acaactgatt	1860
ttgaacatac	cattcgagat	aatgctagtt	actctataac	agtagatttg	agtcaatctt	1920
atthaactat	cccaactaat	caaggaaaat	aacttatgaa	acttcttact	aaagaacggt	1980
ttgatgattt	tcaacacttt	tggtaccaga	tcaatttatt	acaagagagt	aacttcggag	2040
cgttttttga	ccatgataat	aaaaacattc	cacaggttgt	tgcaactatt	gttgatgatt	2100
tacaaggttc	cggaagtctg	aatcatttct	ggatttttgg	caatactact	gatacttcca	2160
tccttatgat	tgctcattta	aatcgaaaat	tctatattca	ggttaattta	aaggactttg	2220
actttgcact	caatttaata	gctataaata	attggaagag	tctcctccaa	actcaacttg	2280
aagctctaaa	cgatacccta	gcaatatttc	aataaataag	gtagaatgga	gtgacaaagc	2340
aacgcgaggg	agactgatta	atgtcatctt	attggaataa	ctatcctgaa	cttaaaaaaa	2400
atattgatga	aaccaatcaa	ctaattcaag	aaagaataca	ggtcagaaat	aaagatattg	2460
aagcggcgct	aagccaactc	acagctgcgg	gaggaaaaca	gctcagacca	gcattctttt	2520
accttttttc	tcaacttgg	aataaggaga	atcaagatac	tcagcaacta	aagaaaatcg	2580
ctgcttcttt	agaaatcctt	cacgttgcta	cattaatcca	tgatgatgtc	attgatgact	2640
caccactaag	acgtggaat	atgaccattc	aaagcaagtt	tggcaaagac	atcgcagttt	2700
atactgggga	tttacttttc	acagtctttt	togatcttat	tttagaatct	atgactgata	2760
caccattttat	gaggattaat	gcaaaaatcta	tgcgtaaaaat	tctcatggga	gaattggacc	2820
agatgcacct	tcgttacaat	caacaacaag	gtatccatca	ctatttacgt	gcgatttcag	2880
gtaagacagc	cgaactcttt	aaattagcta	gcaaagaagg	agcttacttt	ggtgggtgcag	2940
agaaggaggt	tgttcgtcta	gcaggccata	tcggctttaa	cattggtagt	acattccaaa	3000
ttttggatga	tatcctggat	tatactgcag	ataaaaaaac	atttaataag	cctgtcttag	3060
aggatttaac	acaaggcggt	tacagccttc	ctctacttct	tgccattgaa	gaaaatcctg	3120
atattttcaa	acctatttta	gataaaaaaa	cagatatggc	tactgaagac	atggaaaaaa	3180
ttgcttatct	cgctgcttcc	catagaggtg	tgacaaaagc	tcgccatcta	gctcgtaaat	3240
ttactgagaa	agctattagt	gacataaata	agctacccca	gaactctgca	aaaaaacagt	3300
tgctacaatt	aactaattac	cttttaaaaac	gcaaaaattta	aataataaaa	aaacattcca	3360
caatgctaga	aaagcagtta	gggaatggtt	ttttattatc	atthatttat	cgcacctatc	3420
aatcatcata	gatcaccatc	atcagcggct	ttcagctgac	ggtaacgttg	actactttga	3480
gacaattctt	gaggagaacc	ttccaactct	aattgcccac	tttctataaa	taagatacga	3540
tcagcatggt	caataccttt	taagtgatgt	gtaatccaaa	ctaaggctct	accttccaat	3600


```

tctttcataa atacccttag taaggcttgt tcagtaatag gatcaagtcc aacagttggc 3660
tcatctaaga taacaattgg gacatctttt agtaagattc tagccaaagc aattctatgc 3720
ctttcgccac ctgaaaacct aagtccagct tcatcaacca ttgtatagag accatctgat 3780
aaatcagtga ccatctcttt caatccaact cgttcaagaa ctttccatac atcttcttca 3840
ctagcatctt ggtttccaat gcgaatgtta tttagcaggg ttgtattaaa aaggtagggc 3900
gcttggtgta tcaactccaat atagttagaa atgcaatcac caactattga aacatcagca 3960
ccgcctaggg taatcttccc ttgacttggc ttcaagtcgc cacgaagtag actagctaag 4020
gtactcttgc cagaaccact ccgccctaaa atagcaatct tttctccttc ttttaatatcc 4080
aaatctaaat gatgcaaac ccatttctct tgtggcttat actggaaact taaattcttg 4140
acggaaaaat catatggctt attaggcaat t 4171
    
```

```

<210> 33
<211> 649
<212> PRT
<213> Streptococcus
    
```

```

<400> 33
Tyr Asp Asn Ile Phe Gln Ser Leu His His Leu Leu Ala Cys Arg Gly
 1          5          10          15
Lys Ser Gly Asn Thr Leu Ile Asp Gln Leu Val Ala Asp Gly Leu Leu
 20          25          30
His Ala Asp Asn His Tyr His Phe Asn Gly Lys Ser Leu Ala Thr
 35          40          45
Phe Asn Thr Asn Gln Leu Ile Arg Glu Val Val Tyr Val Glu Ile Ser
 50          55          60
Leu Asp Thr Met Ser Ser Gly Glu His Asp Leu Val Lys Val Asn Ile
 65          70          75          80
Ile Arg Pro Thr Thr Glu His Thr Ile Pro Thr Met Met Thr Ala Ser
 85          90          95
Pro Tyr His Gln Gly Ile Asn Asp Pro Ala Ala Asp Gln Lys Thr Tyr
100          105          110
Gln Met Glu Gly Ala Leu Ala Val Lys Gln Pro Lys His Ile Gln Val
115          120          125
Asp Thr Lys Pro Phe Lys Glu Glu Val Lys His Pro Ser Lys Leu Pro
130          135          140
Ile Ser Pro Ala Thr Glu Ser Phe Thr His Ile Asp Ser Tyr Ser Leu
145          150          155          160
Asn Asp Tyr Phe Leu Ser Arg Gly Phe Ala Asn Ile Tyr Val Ser Gly
165          170          175
Val Gly Thr Ala Gly Ser Thr Gly Phe Met Thr Ser Gly Asp Tyr Gln
180          185          190
Gln Ile Gln Ser Phe Lys Ala Val Ile Asp Trp Leu Asn Gly Lys Val
195          200          205
Thr Ala Phe Thr Ser His Lys Arg Asp Lys Gln Val Lys Ala Asp Trp
210          215          220
Ser Asn Gly Leu Val Ala Thr Thr Gly Lys Ser Tyr Leu Gly Thr Met
225          230          235          240
Ser Thr Gly Leu Ala Thr Thr Gly Val Glu Gly Leu Lys Val Ile Ile
245          250          255
Ala Glu Ala Ala Ile Ser Thr Trp Tyr Asp Tyr Tyr Arg Glu Asn Gly
260          265          270
Leu Val Cys Ser Pro Gly Gly Tyr Pro Gly Glu Asp Leu Asp Val Leu
275          280          285
Thr Glu Leu Thr Tyr Ser Arg Asn Leu Leu Ala Gly Asp Tyr Ile Lys
290          295          300
Asn Asn Asp Cys Tyr Gln Ala Leu Leu Asn Glu Gln Ser Lys Ala Ile
    
```

305 310 315 320
 Asp Arg Gln Ser Gly Asp Tyr Asn Gln Tyr Trp His Asp Arg Asn Tyr
 325 330 335
 Leu Thr His Val Asn Asn Val Lys Ser Arg Val Val Tyr Thr His Gly
 340 345 350
 Leu Gln Asp Trp Asn Val Lys Pro Arg His Val Tyr Lys Val Phe Asn
 355 360 365
 Ala Leu Pro Gln Thr Ile Lys Lys His Leu Phe Leu His Gln Gly Gln
 370 375 380
 His Val Tyr Met His Asn Trp Gln Ser Ile Asp Phe Arg Glu Ser Met
 385 390 395 400
 Asn Ala Leu Leu Ser Gln Glu Leu Leu Gly Ile Asp Asn His Phe Gln
 405 410 415
 Leu Glu Glu Val Ile Trp Gln Asp Asn Thr Thr Glu Gln Thr Trp Gln
 420 425 430
 Val Leu Asp Ala Phe Gly Gly Asn His Gln Glu Gln Ile Gly Leu Gly
 435 440 445
 Asp Ser Lys Lys Leu Ile Asp Asn His Tyr Asp Lys Glu Ala Phe Asp
 450 455 460
 Thr Tyr Cys Lys Asp Phe Asn Val Phe Lys Asn Asp Leu Phe Lys Gly
 465 470 475 480
 Asn Asn Lys Thr Asn Gln Ile Thr Ile Asn Leu Pro Leu Lys Lys Asn
 485 490 495
 Tyr Leu Leu Asn Gly Gln Cys Lys Leu His Leu Arg Val Lys Thr Ser
 500 505 510
 Asp Lys Lys Ala Ile Leu Ser Ala Gln Ile Leu Asp Tyr Gly Pro Lys
 515 520 525
 Lys Arg Phe Lys Asp Thr Pro Thr Ile Lys Phe Leu Asn Ser Leu Asp
 530 535 540
 Asn Gly Lys Asn Phe Ala Arg Glu Ala Leu Arg Glu Leu Pro Phe Thr
 545 550 555 560
 Lys Asp His Tyr Arg Val Ile Ser Lys Gly Val Leu Asn Leu Gln Asn
 565 570 575
 Arg Thr Asp Leu Leu Thr Ile Glu Ala Ile Glu Pro Glu Gln Trp Phe
 580 585 590
 Asp Ile Glu Phe Ser Leu Gln Pro Ser Ile Tyr Gln Leu Ser Lys Gly
 595 600 605
 Asp Asn Leu Arg Ile Ile Leu Tyr Thr Thr Asp Phe Glu His Thr Ile
 610 615 620
 Arg Asp Asn Ala Ser Tyr Ser Ile Thr Val Asp Leu Ser Gln Ser Tyr
 625 630 635 640
 Leu Thr Ile Pro Thr Asn Gln Gly Asn
 645

<210> 34
 <211> 119
 <212> PRT
 <213> Streptococcus

<400> 34
 Met Lys Leu Leu Thr Lys Glu Arg Phe Asp Asp Ser Gln His Phe Trp
 1 5 10 15
 Tyr Gln Ile Asn Leu Leu Gln Glu Ser Asn Phe Gly Ala Val Phe Asp
 20 25 30
 His Asp Asn Lys Asn Ile Pro Gln Val Val Ala Thr Ile Val Asp Asp
 35 40 45

Leu Gln Gly Ser Gly Ser Ser Asn His Phe Trp Tyr Phe Gly Asn Thr
 50 55 60
 Thr Asp Thr Ser Ile Leu Met Ile Ala His Leu Asn Arg Lys Phe Tyr
 65 70 75 80
 Ile Gln Val Asn Leu Lys Asp Phe Asp Phe Ala Leu Asn Leu Ile Ala
 85 90 95
 Ile Asn Asn Trp Lys Ser Leu Leu Gln Thr Gln Leu Glu Ala Leu Asn
 100 105 110
 Asp Thr Leu Ala Ile Phe Gln
 115

<210> 35
 <211> 326
 <212> PRT
 <213> Streptococcus

<400> 35
 Met Ser Ser Tyr Trp Asn Asn Tyr Pro Glu Leu Lys Lys Asn Ile Asp
 1 5 10 15
 Glu Thr Asn Gln Leu Ile Gln Glu Arg Ile Gln Val Arg Asn Lys Asp
 20 25 30
 Ile Glu Ala Ala Leu Ser Gln Leu Thr Ala Ala Gly Gly Lys Gln Leu
 35 40 45
 Arg Pro Ala Phe Phe Tyr Leu Phe Ser Gln Leu Gly Asn Lys Glu Asn
 50 55 60
 Gln Asp Thr Gln Gln Leu Lys Lys Ile Ala Ala Ser Leu Glu Ile Leu
 65 70 75 80
 His Val Ala Thr Leu Ile His Asp Asp Val Ile Asp Asp Ser Pro Leu
 85 90 95
 Arg Arg Gly Asn Met Thr Ile Gln Ser Lys Phe Gly Lys Asp Ile Ala
 100 105 110
 Val Tyr Thr Gly Asp Leu Leu Phe Thr Val Phe Phe Asp Leu Ile Leu
 115 120 125
 Glu Ser Met Thr Asp Thr Pro Phe Met Arg Ile Asn Ala Lys Ser Met
 130 135 140
 Arg Lys Ile Leu Met Gly Glu Leu Asp Gln Met His Leu Arg Tyr Asn
 145 150 155 160
 Gln Gln Gln Gly Ile His His Tyr Leu Arg Ala Ile Ser Gly Lys Thr
 165 170 175
 Ala Glu Leu Phe Lys Leu Ala Ser Lys Glu Gly Ala Tyr Phe Gly Gly
 180 185 190
 Ala Glu Lys Glu Val Val Arg Leu Ala Gly His Ile Gly Phe Asn Ile
 195 200 205
 Gly Met Thr Phe Gln Ile Leu Asp Asp Ile Leu Asp Tyr Thr Ala Asp
 210 215 220
 Lys Lys Thr Phe Asn Lys Pro Val Leu Glu Asp Leu Thr Gln Gly Val
 225 230 235 240
 Tyr Ser Leu Pro Leu Leu Leu Ala Ile Glu Glu Asn Pro Asp Ile Phe
 245 250 255
 Lys Pro Ile Leu Asp Lys Lys Thr Asp Met Ala Thr Glu Asp Met Glu
 260 265 270
 Lys Ile Ala Tyr Leu Val Val Ser His Arg Gly Val Asp Lys Ala Arg
 275 280 285
 His Leu Ala Arg Lys Phe Thr Glu Lys Ala Ile Ser Asp Ile Asn Lys
 290 295 300
 Leu Pro Gln Asn Ser Ala Lys Lys Gln Leu Leu Gln Leu Thr Asn Tyr

305						310					315					320
Leu	Leu	Lys	Arg	Lys	Ile											
				325												

<210> 36
 <211> 247
 <212> PRT
 <213> Streptococcus

<400> 36
 Leu Pro Asn Lys Pro Tyr Asp Phe Ser Val Lys Asn Leu Ser Phe Gln
 1 5 10 15
 Tyr Lys Pro Gln Glu Lys Trp Val Leu His His Leu Asp Leu Asp Ile
 20 25 30
 Lys Glu Gly Glu Lys Ile Ala Ile Leu Gly Arg Ser Gly Ser Gly Lys
 35 40 45
 Ser Thr Leu Ala Ser Leu Leu Arg Gly Asp Leu Lys Ala Ser Gln Gly
 50 55 60
 Lys Ile Thr Leu Gly Gly Ala Asp Val Ser Ile Val Gly Asp Cys Ile
 65 70 75 80
 Ser Asn Tyr Ile Gly Val Ile Gln Gln Ala Pro Tyr Leu Phe Asn Thr
 85 90 95
 Thr Leu Leu Asn Asn Ile Arg Ile Gly Asn Gln Asp Ala Ser Glu Glu
 100 105 110
 Asp Val Trp Lys Val Leu Glu Arg Val Gly Leu Lys Glu Met Val Thr
 115 120 125
 Asp Leu Ser Asp Gly Leu Tyr Thr Met Val Asp Glu Ala Gly Leu Arg
 130 135 140
 Phe Ser Gly Gly Glu Arg His Arg Ile Ala Leu Ala Arg Ile Leu Leu
 145 150 155 160
 Lys Asp Val Pro Ile Val Ile Leu Asp Glu Pro Thr Val Gly Leu Asp
 165 170 175
 Pro Ile Thr Glu Gln Ala Leu Leu Arg Val Phe Met Lys Glu Leu Glu
 180 185 190
 Gly Lys Thr Leu Val Trp Ile Thr His His Leu Lys Gly Ile Glu His
 195 200 205
 Ala Asp Arg Ile Leu Phe Ile Glu Asn Gly Gln Leu Glu Leu Glu Gly
 210 215 220
 Ser Pro Gln Glu Leu Ser Gln Ser Ser Gln Arg Tyr Arg Gln Leu Lys
 225 230 235 240
 Ala Ala Asp Asp Gly Asp Leu
 245

<210> 37
 <211> 3480
 <212> DNA
 <213> Streptococcus

<400> 37
 aattctatattt ggagggtttttt cttgaataaaa tggtttagtta aggcaagttc cttagttggt 60
 ttaggtggta tggttttatc tgcgggttcc cgagtttttag cggataactta tgtccgtcca 120
 attgataatg gtagaattac aacagggtttc aatggttatc ctggacattg tgggggtggat 180
 tatgctgttc cgactggaac gattattagg gcagtggcag atggtactgt gaaatttga 240
 ggagctggag ccaacttttc ttggatgaca gacttagcag gaaattgtgt catgattcaa 300
 catgctggat gaatgcgatg tggttacgct catatgtcac gtgtggtggc taggactggg 360
 gaaaaagtca aacaaggaga tatcatcggt tacgtaggag caactggtat ggcgacggga 420

cctcaccttc attttgaatt tttaccagct aaccctaatt ttcaaaatgg tttccatgga 480
 cgtatcaatc caacgtcact aattgctaac gttgcgacct ttagtggaaa aacgcaagca 540
 tcagctccaa gcattaagcc attacaatca gtcctgttac agaatcaatc tagtaaatta 600
 aaagtgtatc gagtagatga attacaaaag gttaatgggtg tttggttagt caaaaataac 660
 accctaacgc cgactgggtt tgattggaac gataatggta taccagcatc agaaattgat 720
 gaggttgaatg ctaatggtaa tttgacagct gaccagggttc ttcaaaaagg tggttacttt 780
 atctttaatc ctaaaactct taagactgta gaaaaaccca tccaaggaac agctggttta 840
 acttgggcta agacacgctt tgctaattgg agttcagttt ggcttcgctg tgacaacagt 900
 caagaactgc tttacaaata gtttgaggta ttgattcatt gttttaaatg acagttttgt 960
 tactaaactaa gtacaatttc tttaaaccgt ctgaaaataa ttttatagtc cagtaaagtg 1020
 tgatattata gtctcggact aataaaaagg aaataggaat tgaagcaatg aaaatgaata 1080
 aaaaggtact attgacatcg acaatggcag ctctcgctatt atcagtcgca agtgttcaag 1140
 cacaagaaac agatacgacg tggacagcac gtactgtttc agaggtaaag gctgatttgg 1200
 taaagcaaga caataaatca tcatatactg tgaaatatgg tgatacacta agcgttattt 1260
 cagaagcaat gtcaattgat atgaatgtct tagcaaaaa taataacatt gcagatatca 1320
 atcttattta tcctgagaca acactgacag taacttacga tcagaagagt catactgcca 1380
 ctccaatgaa aatagaaaca ccagcaacaa atgctgctgg tcaacaaca gctactgtgg 1440
 atttgaaaac caatcaagtt tctgttgacg accaaaaagt ttctctcaat acaatttcgg 1500
 aaggtatgac accagaagca gcaacaacga ttgtttcgcc aatgaagaca tattcttctg 1560
 cgccagcttt gaaatcaaaa gaagtattag cacaagagca agctgttagt caagcagcag 1620
 ctaatgaaca ggtatcaaca gtcctgtgta agtcgattac ttcagaagtt ccagcagcta 1680
 aagaggaagt taaaccaact cagacgtcag tcagtcagtc aacaacagta tcaccagctt 1740
 ctgttgccgc tgaaacacca gctccagtag ctaaagttag accggtaaga actgtagcag 1800
 ccctaagat ggcaagtgtt aaagttagca ctccataagt agaaactggg gcattaccag 1860
 agcatgtatc agctccagca gttcctgtga ctacgacttc aacagctaca gacagtaagt 1920
 tacaagcgac tgaagttaag agcgttccgg tagcacaaaa agctccaaca gcaacaccgg 1980
 tagcacaacc agcttcaaca acaaatgcag tagctgcaca tcctgaaaat gcagggctcc 2040
 aacctcatgt tgcagcttat aaagaaaaag tagcgtcaac ttatggagtt aatgaattca 2100
 gtacataccg tgcaggtgat ccaggtgatc atggtaaagg ttttagcagtc gactttattg 2160
 taggtaaaaa ccaagcactt ggtaatgaag ttgcacagta ctctacaca aatattggcag 2220
 caaataacat ttcatatggt atctggcaac aaaagtttta ctcaaataca aatagtattt 2280
 atggacctgc taatacttgg aatgcaatgc cagatcgtgg tggcgttact gccaaccatt 2340
 atgacctatg tcacgtatca tttaaacaa atataaaaa aggaagctat ttggcttctt 2400
 ttttatatgc cttgaaataga ctttcaaggt tcttatctaa tttttattaa attgaggaga 2460
 ttaagctata agtctgaaac tactttcacg ttaaccgtga ctaaatacaa acgttaaaac 2520
 taaaatctaa gtctgtaaaag attattgaaa acgctttaaa aacagatata ataaggtttg 2580
 tagatatcta aaattaaaaa agataaggaa gtgagaatat gccacatcta agtaaagaag 2640
 cttttaaaaa gcaataaaaa aatggcatta ttgtgtcatg tcaagctttg cctggggagc 2700
 ctctttatac tgaaagtgga ggtgttatgc ctcttttagc tttggcagct caagaagcag 2760
 gagcggttgg tataagagcc aatagtgtcc gcgacattaa ggaaattcaa gaagttacta 2820
 atttacctat catcggcatt attaaacgtg aatatacctcc acaagaacca tttatcactg 2880
 ctacgatgac agaggtggat caattagcta gtttagatat tgcagtaata gccttagatt 2940
 gtacacttag agagcgtcat gatggtttga gtgtagctga gtttattcaa aagataaaaag 3000
 ggaaatatcc tgaacagttg ctaatggctg atataagtac ttttgaagaa ggtaaaaatg 3060
 cttttgaagc aggagttgat tttgtgggta caactctatc tggatacaca gattacagcc 3120
 gccaagaaga aggaccggat atagaactcc ttaataagct ttgtcaagcc ggtatagatg 3180
 tgattgcgga aggtaaaatt catactocta agcaagctaa tgaaattaat catataggtg 3240
 ttgcaggaat tgtagtgggt ggtgctatca ctagacccaaa agaaatagcg gagcgtttca 3300
 tctcaggact tagttaaag tgttactcaa aaatcaaaa caaaaataaaa aaggggaata 3360
 gttatgagta tcaaaaaaag tgtgattgggt ttttgctcag gagctgcagc attatcaatg 3420
 tttgcttgtg tagacagtag tcaatctggt atggctgccc agaaggataa agtcgaaatt 3480

- <210> 38
- <211> 306
- <212> PRT
- <213> Streptococcus

<400> 38
 Asn Ser Ile Trp Arg Phe Phe Leu Asn Lys Trp Leu Val Lys Ala Ser
 1 5 10 15
 Ser Leu Val Val Leu Gly Gly Met Val Leu Ser Ala Gly Ser Arg Val
 20 25 30
 Leu Ala Asp Thr Tyr Val Arg Pro Ile Asp Asn Gly Arg Ile Thr Thr
 35 40 45
 Gly Phe Asn Gly Tyr Pro Gly His Cys Gly Val Asp Tyr Ala Val Pro
 50 55 60
 Thr Gly Thr Ile Ile Arg Ala Val Ala Asp Gly Thr Val Lys Phe Ala
 65 70 75 80
 Gly Ala Gly Ala Asn Phe Ser Trp Met Thr Asp Leu Ala Gly Asn Cys
 85 90 95
 Val Met Ile Gln His Ala Asp Gly Met His Ser Gly Tyr Ala His Met
 100 105 110
 Ser Arg Val Val Ala Arg Thr Gly Glu Lys Val Lys Gln Gly Asp Ile
 115 120 125
 Ile Gly Tyr Val Gly Ala Thr Gly Met Ala Thr Gly Pro His Leu His
 130 135 140
 Phe Glu Phe Leu Pro Ala Asn Pro Asn Phe Gln Asn Gly Phe His Gly
 145 150 155 160
 Arg Ile Asn Pro Thr Ser Leu Ile Ala Asn Val Ala Thr Phe Ser Gly
 165 170 175
 Lys Thr Gln Ala Ser Ala Pro Ser Ile Lys Pro Leu Gln Ser Ala Pro
 180 185 190
 Val Gln Asn Gln Ser Ser Lys Leu Lys Val Tyr Arg Val Asp Glu Leu
 195 200 205
 Gln Lys Val Asn Gly Val Trp Leu Val Lys Asn Asn Thr Leu Thr Pro
 210 215 220
 Thr Gly Phe Asp Trp Asn Asp Asn Gly Ile Pro Ala Ser Glu Ile Asp
 225 230 235 240
 Glu Val Asp Ala Asn Gly Asn Leu Thr Ala Asp Gln Val Leu Gln Lys
 245 250 255
 Gly Gly Tyr Phe Ile Phe Asn Pro Lys Thr Leu Lys Thr Val Glu Lys
 260 265 270
 Pro Ile Gln Gly Thr Ala Gly Leu Thr Trp Ala Lys Thr Arg Phe Ala
 275 280 285
 Asn Gly Ser Ser Val Trp Leu Arg Val Asp Asn Ser Gln Glu Leu Leu
 290 295 300
 Tyr Lys
 305

<210> 39
 <211> 434
 <212> PRT
 <213> Streptococcus

<400> 39
 Met Lys Met Asn Lys Lys Val Leu Leu Thr Ser Thr Met Ala Ala Ser
 1 5 10 15
 Leu Leu Ser Val Ala Ser Val Gln Ala Gln Glu Thr Asp Thr Thr Trp
 20 25 30
 Thr Ala Arg Thr Val Ser Glu Val Lys Ala Asp Leu Val Lys Gln Asp
 35 40 45
 Asn Lys Ser Ser Tyr Thr Val Lys Tyr Gly Asp Thr Leu Ser Val Ile

50						55						60					
Ser	Glu	Ala	Met	Ser	Ile	Asp	Met	Asn	Val	Leu	Ala	Lys	Ile	Asn	Asn		
65						70						75			80		
Ile	Ala	Asp	Ile	Asn	Leu	Ile	Tyr	Pro	Glu	Thr	Thr	Leu	Thr	Val	Thr		
				85					90					95			
Tyr	Asp	Gln	Lys	Ser	His	Thr	Ala	Thr	Ser	Met	Lys	Ile	Glu	Thr	Pro		
			100					105						110			
Ala	Thr	Asn	Ala	Ala	Gly	Gln	Thr	Thr	Ala	Thr	Val	Asp	Leu	Lys	Thr		
			115					120						125			
Asn	Gln	Val	Ser	Val	Ala	Asp	Gln	Lys	Val	Ser	Leu	Asn	Thr	Ile	Ser		
						135								140			
Glu	Gly	Met	Thr	Pro	Glu	Ala	Ala	Thr	Thr	Ile	Val	Ser	Pro	Met	Lys		
145						150								155			160
Thr	Tyr	Ser	Ser	Ala	Pro	Ala	Leu	Lys	Ser	Lys	Glu	Val	Leu	Ala	Gln		
				165						170							175
Glu	Gln	Ala	Val	Ser	Gln	Ala	Ala	Ala	Asn	Glu	Gln	Val	Ser	Thr	Ala		
				180						185							190
Pro	Val	Lys	Ser	Ile	Thr	Ser	Glu	Val	Pro	Ala	Ala	Lys	Glu	Glu	Val		
										200							205
Lys	Pro	Thr	Gln	Thr	Ser	Val	Ser	Gln	Ser	Thr	Thr	Val	Ser	Pro	Ala		
														210			
Ser	Val	Ala	Ala	Glu	Thr	Pro	Ala	Pro	Val	Ala	Lys	Val	Ala	Pro	Val		
225						230											240
Arg	Thr	Val	Ala	Ala	Pro	Arg	Val	Ala	Ser	Val	Lys	Val	Val	Thr	Pro		
						245											255
Lys	Val	Glu	Thr	Gly	Ala	Ser	Pro	Glu	His	Val	Ser	Ala	Pro	Ala	Val		
						260											270
Pro	Val	Thr	Thr	Thr	Ser	Thr	Ala	Thr	Asp	Ser	Lys	Leu	Gln	Ala	Thr		
																	285
Glu	Val	Lys	Ser	Val	Pro	Val	Ala	Gln	Lys	Ala	Pro	Thr	Ala	Thr	Pro		
																	300
Val	Ala	Gln	Pro	Ala	Ser	Thr	Thr	Asn	Ala	Val	Ala	Ala	His	Pro	Glu		
305						310											320
Asn	Ala	Gly	Leu	Gln	Pro	His	Val	Ala	Ala	Tyr	Lys	Glu	Lys	Val	Ala		
						325											335
Ser	Thr	Tyr	Gly	Val	Asn	Glu	Phe	Ser	Thr	Tyr	Arg	Ala	Gly	Asp	Pro		
																	350
Gly	Asp	His	Gly	Lys	Gly	Leu	Ala	Val	Asp	Phe	Ile	Val	Gly	Lys	Asn		
																	365
Gln	Ala	Leu	Gly	Asn	Glu	Val	Ala	Gln	Tyr	Ser	Thr	Gln	Asn	Met	Ala		
																	380
Ala	Asn	Asn	Ile	Ser	Tyr	Val	Ile	Trp	Gln	Gln	Lys	Phe	Tyr	Ser	Asn		
385						390											400
Thr	Asn	Ser	Ile	Tyr	Gly	Pro	Ala	Asn	Thr	Trp	Asn	Ala	Met	Pro	Asp		
																	415
Arg	Gly	Gly	Val	Thr	Ala	Asn	His	Tyr	Asp	His	Val	His	Val	Ser	Phe		
																	430
Asn	Lys																

- <210> 40
- <211> 232
- <212> PRT
- <213> Streptococcus
- <400> 40

Met Pro His Leu Ser Lys Glu Ala Phe Lys Lys Gln Ile Lys Asn Gly
 1 5 10 15
 Ile Ile Val Ser Cys Gln Ala Leu Pro Gly Glu Pro Leu Tyr Thr Glu
 20 25 30
 Ser Gly Gly Val Met Pro Leu Leu Ala Leu Ala Ala Gln Glu Ala Gly
 35 40 45
 Ala Val Gly Ile Arg Ala Asn Ser Val Arg Asp Ile Lys Glu Ile Gln
 50 55 60
 Glu Val Thr Asn Leu Pro Ile Ile Gly Ile Ile Lys Arg Glu Tyr Pro
 65 70 75 80
 Pro Gln Glu Pro Phe Ile Thr Ala Thr Met Thr Glu Val Asp Gln Leu
 85 90 95
 Ala Ser Leu Asp Ile Ala Val Ile Ala Leu Asp Cys Thr Leu Arg Glu
 100 105 110
 Arg His Asp Gly Leu Ser Val Ala Glu Phe Ile Gln Lys Ile Lys Gly
 115 120 125
 Lys Tyr Pro Glu Gln Leu Leu Met Ala Asp Ile Ser Thr Phe Glu Glu
 130 135 140
 Gly Lys Asn Ala Phe Glu Ala Gly Val Asp Phe Val Gly Thr Thr Leu
 145 150 155 160
 Ser Gly Tyr Thr Asp Tyr Xaa Arg Gln Glu Glu Gly Pro Asp Ile Glu
 165 170 175
 Leu Leu Asn Lys Leu Cys Gln Ala Gly Ile Asp Val Ile Ala Glu Gly
 180 185 190
 Lys Ile His Thr Pro Lys Gln Ala Asn Glu Ile Asn His Ile Gly Val
 195 200 205
 Ala Gly Ile Val Val Gly Gly Ala Ile Thr Arg Pro Lys Glu Ile Ala
 210 215 220
 Glu Arg Phe Ile Ser Gly Leu Ser
 225 230

<210> 41
 <211> 39
 <212> PRT
 <213> Streptococcus

<400> 41
 Met Ser Ile Lys Lys Ser Val Ile Gly Phe Cys Leu Gly Ala Ala Ala
 1 5 10 15
 Leu Ser Met Phe Ala Cys Val Asp Ser Ser Gln Ser Val Met Ala Ala
 20 25 30
 Glu Lys Asp Lys Val Glu Ile
 35

<210> 42
 <211> 1305
 <212> DNA
 <213> Streptococcus

<400> 42
 atgaaaatga ataaaaagggt actattgaca tcgacaatgg cagcttcgct attatcagtc 60
 gcaagtgttc aagcacaaga aacagatacg acgtggacag cacgtactgt ttcagaggta 120
 aaggctgatt tggtaaagca agacaataaa tcatcatata ctgtgaaata tgggtataca 180
 ctaagcgtta tttcagaagc aatgtcaatt gatatgaatg tcttagcaaa aattaataac 240
 attgcagata tcaatcttat ttatcctgag acaacactga cagtaactta cgatcagaag 300
 agtcatactg ccacttcaat gaaaatagaa acaccagcaa caaatgctgc tgggtcaaca 360


```

acagctactg tggatttgaa aaccaatcaa gtttctggtg cagaccacaaa agtttctctc 420
aatacaattht cgggaaggat gacaccagaa gcagcaacaa cgattgtttc gccaatgaag 480
acataattctt ctgcgccagc tttgaaatca aaagaagat tagcacaaga gcaagctggt 540
agtcaagcag cagctaataga acaggtatca acagctcctg tgaagtcgat tacttcagaa 600
gttccagcag ctaaagagga agttaaacca actcagacgt cagtcagtcg gtcaacaaca 660
gtatcaccag cttctggtgc cgctgaaaca ccagctccag tagctaaagt agcaccggta 720
agaactgtag cagcccctag agtggcaagt gttaaagtag tcactcctaa agtagaaact 780
ggtgcatcac cagagcatgt atcagctcca gcagttcctg tgactacgac ttcaacagct 840
acagacagta agttacaagc gactgaagtt aagagcgttc cggtagcaca aaaagctcca 900
acagcaacac cggtagcaca accagcttca acaaacaaatg cagtagctgc acatcctgaa 960
aatgcagggc tccaacctca tgttgcagct tataaagaaa aagtagcgtc aacttatgga 1020
gttaatgaat tcagtacata ccgtgcaggt gatccaggtg atcatggtaa aggttttagca 1080
gtcgcacttta ttgtaggtaa aaaccaagca cttggtaatg aagttgcaca gtactctaca 1140
caaaatatgg cagcaataaa catttcatat gttatctggc aacaaaagtt ttactcaaat 1200
acaaatagta tttatggacc tgctaatact tggaatgcaa tgccagatcg tggtagcgtt 1260
actgccaacc attatgacca tgttcacgta tcatttaaca aataa 1305

```

```

<210> 43
<211> 1230
<212> DNA
<213> Streptococcus

```

```

<400> 43
caagaaacag atacgacgtg gacagcacgt actgtttcag aggtaaaggc tgatttggtg 60
aagcaagaca ataaatcatc atatactgtg aaatatgggtg atacactaag cgttattttca 120
gaagcaatgt caattgatat gaatgtctta gcaaaaatta ataacattgc agatatcaat 180
cttattttatc ctgagacaac actgacagta acttacgatc agaagagtca tactgccact 240
tcaatgaaaa tagaaacacc agcaacaaat gctgctgggtc aaacaacagc tactgtggat 300
ttgaaaacca atcaagtttc tgttgcagac caaaaagttt ctctcaatac aatttcggaa 360
ggtatgacac cagaagcagc aacaacgatt gtttcgccaa tgaagacata ttcttctgcg 420
ccagctttga aatcaaaaga agtattagca caagagcaag ctgtagtca agcagcagct 480
aatgaacagg tatcaacagc tcctgtgaag tgcattactt cagaagttcc agcagctaaa 540
gaggaagtta aaccaactca gacgtcagtc agtcagtcaa caacagtatc accagcttct 600
gttgccgctg aaacaccagc tccagtagct aaagtagcac cggtaagaac tgtagcagcc 660
cctagagtggt caagtgttaa agtagtcact cctaaagtag aaactgggtg atcaccagag 720
catgtatcag ctccagcagt tcctgtgact acgacttcaa cagctacaga cagtaagtta 780
caagcgactg aagttaagag cgttccggta gcacaaaaag ctccaacagc aacaccggta 840
gcacaaccag cttcaacaac aaatgcagta gctgcacatc ctgaaaatgc agggctccaa 900
cctcatgttg cagcttataa agaaaaagta gcgtcaactt atggagttaa tgaattcagt 960
acataaccgtg caggtgatcc aggtgatcat ggtaaagggt tagcagtcga ctttattgta 1020
ggtaaaaacc aagcacttgg taatgaagtt gcacagtact ctacacaaaa tatggcagca 1080
aataacattt catatggtat ctggcaacaa aagttttact caaatacaaa tagtattttat 1140
ggacctgcta atacttggaa tgcaatgcca gatcgtgggtg gcgttactgc caaccattat 1200
gaccatgttc acgtatcatt taacaaataa

```

```

<210> 44
<211> 409
<212> PRT
<213> Streptococcus

```

```

<400> 44
Gln Glu Thr Asp Thr Trp Thr Ala Arg Thr Val Ser Glu Val Lys
1 5 10 15
Ala Asp Leu Val Lys Gln Asp Asn Lys Ser Ser Tyr Thr Val Lys Tyr
20 25 30
Gly Asp Thr Leu Ser Val Ile Ser Glu Ala Met Ser Ile Asp Met Asn

```




INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification⁶ : C12N 15/31, C07K 14/315, A61K 39/09, C12N 1/21	A3	(11) International Publication Number: WO 99/42588 (43) International Publication Date: 26 August 1999 (26.08.99)
(21) International Application Number: PCT/CA99/00114 (22) International Filing Date: 17 February 1999 (17.02.99) (30) Priority Data: 60/075,425 20 February 1998 (20.02.98) US (71) Applicant (for all designated States except US): BIOCHEM VACCINS INC. [CA/CA]; 2323 boulevard du Parc Technologique, Sainte-Foy, Québec G1P 4R8 (CA). (72) Inventors; and (75) Inventors/Applicants (for US only): BRODEUR, Bernard, R. [CA/CA]; 2401 rue Maritain, Sillery, Québec G1T 1N6 (CA). RIOUX, Clément [CA/CA]; 1012 Jean-Charles Cantin, Ville de Cap Rouge, Québec G1Y 2X1 (CA). BOYER, Martine [CA/CA]; Apt. 204, 25 des Mouettes, Beauport, Québec G1E 7G1 (CA). CHARLEBOIS, Isabelle [CA/CA]; 410 Mirabel, St-Nicolas, Québec G7A 2L5 (CA). HAMEL, Josée [CA/CA]; 2401 rue Maritain, Sillery, Québec G1T 1N6 (CA). MARTIN, Denis [CA/CA]; 4728-G rue Gaboury, St-Augustin-de-Desmaures, Québec G3A 1E9 (CA).	(74) Agents: CÔTE, France et al.; Swabey Ogilvy Renault, Suite 1600, 1981 McGill College Avenue, Montréal, Québec H3A 2Y3 (CA). (81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). Published <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i> (88) Date of publication of the international search report: 23 March 2000 (23.03.00)	
(54) Title: GROUP B STREPTOCOCCUS ANTIGENS		
(57) Abstract <p>Group B streptococcus (GBS) proteins and polynucleotides encoding them are disclosed. Said proteins are antigenic and therefore useful vaccine components for the prophylaxis or therapy of streptococcus infection in animals. Also disclosed are recombinant methods of producing the protein antigens as well as diagnostic assays for detecting streptococcus bacterial infection.</p>		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

INTERNATIONAL SEARCH REPORT

International Application No

PCT/CA 99/00114

A. CLASSIFICATION OF SUBJECT MATTER
 IPC 6 C12N15/31 C07K14/315 A61K39/09 C12N1/21

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C07K C12N A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	MICHEL J L ET AL: "Cloned alpha and beta C-protein antigens of group B Streptococci elicit protective immunity" INFECTION AND IMMUNITY., vol. 59, no. 6, June 1991 (1991-06), pages 2023-2028, XP002107260 AMERICAN SOCIETY FOR MICROBIOLOGY. WASHINGTON., US ISSN: 0019-9567 the whole document	1-48
	--- -/--	

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

* Special categories of cited documents :

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- *Z* document member of the same patent family

Date of the actual completion of the international search
 15 December 1999

Date of mailing of the international search report
 24 01 2000

Name and mailing address of the ISA
 European Patent Office, P.B. 5818 Patentlaan 2
 NL - 2280 HV Rijswijk
 Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
 Fax: (+31-70) 340-3016

Authorized officer
 Lejeune, R

INTERNATIONAL SEARCH REPORT

International Application No

PCT/CA 99/00114

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>LACHENAUER C S ET AL: "Cloning and expression in Escherichia coli of a protective surface protein from type V group B Streptococci" ADVANCES IN EXPERIMENTAL MEDICINE AND BIOLOGY, vol. 418, 9 December 1997 (1997-12-09), pages 615-618, XP002107261 SPRING ST., NY, US ISSN: 0065-2598 the whole document</p>	1-48
P,X	<p>--- DATABASE EMBL [Online] Accession number AF062533, 11 February 1999 (1999-02-11) SPELLERBERG B ET AL: "Streptococcus agalactiae Lmb (lmb) gene, complete cds; and unknown gene." XP002125180 98.9% identity between base 1-2514 of SEQ ID NO 13 and base 988-3501 of AF062533 Translation product (AC: Q9ZHG9) has 98.5% identity in 793 AA overlap with SEQ ID NO 15 and 98.5% identity in 715 AA overlap with SEQ ID 16 & SPELLERBERG B ET AL: "Lmb, a protein with similarities to the LraI adhesin family, mediates attachment of Streptococcus agalactiae to human laminin" INFECTION AND IMMUNITY., vol. 67, no. 2, February 1999 (1999-02), pages 871-878, AMERICAN SOCIETY FOR MICROBIOLOGY. WASHINGTON., US ISSN: 0019-9567</p>	1-10, 16-23,26
X	<p>--- DATABASE EMBL [Online] Accession Number L23843, 4 January 1994 (1994-01-04) MACRINA F L ET AL: "ISN IS199 from Streptococcus mutans IS3 (Brathall serotype C) DNA fragment" XP002125181 79.6% identity between base 5212-4314 of SEQ ID NO 13 and base 312-1220 of L23843 Translation has 83.4% identity in 283 AA overlap with SEQ ID NO 21 --- -/--</p>	1,3-7,10

INTERNATIONAL SEARCH REPORT

International Application No

PCT/CA 99/00114

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>DATABASE EMBL [Online] Accession Number AF026542, 15 October 1997 (1997-10-15) HYNES W L ET AL: "Streptococcus pyogenes FF22 lantibiotic (scn) gene cluster region containing: scnK, scnR, streptococin A-FF22 precursor (scnA), scnA1, scnM, scnT, scnF, scnE, scnG genes, complete cds, and tnpA gene, partial cds." XP002125182 88.2% identity between base 2607-2953 of SEQ ID NO 13 and base 10435-10777 of AF026542 Translation product (AC: 031057) has 95.8% identity in 71 AA overlap with SEQ ID NO 17</p>	<p>1-10, 16-23,26</p>
P,X	<p>--- DATABASE GENESEQ [Online] Accession Number V52136, 23 October 1998 (1998-10-23) BARASH S C ET AL: "Streptococcus pneumoniae genome fragment SEQ ID NO:3" XP002125183 68.5% identity between base 2539-3319 of SEQ ID NO 37 and base 18492-19271 of V52136 Translation has 74.5% identity in 231 AA overlap with SEQ ID NO 40 & WO 98 18931 A (DOUGHERTY BRIAN A ;HUMAN GENOME SCIENCES INC (US); ROSEN CRAIG A) 7 May 1998 (1998-05-07) -----</p>	<p>1,3-7,10</p>

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CA 99/00114

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

Although claims 37-46 are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.
2. Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
3. Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

As a result of the prior review under R. 40.2(e) PCT,
no additional fees are to be refunded.

1. As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

11-14,16,24,25,27,28,30,31 (completely), 1-10,15,17-23,26,29,32-48 (all partially) i.e. (group of) inventions 1, 3 and 7
4. No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- The additional search fees were accompanied by the applicant's protest.
- No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 1-10,15,17-23,26,29,32-48 (all partially)

An isolated polynucleotide encoding a polypeptide having a sequence selected from the group consisting of SEQ ID 2, SEQ ID NO 3, SEQ ID NO 4, SEQ ID NO 5, SEQ ID NO 6 i.e. the open reading frames of clone 1 (SEQ ID NO 1). Also a vector comprising the polynucleotide, a host cell transformed therewith, an isolated polypeptide encoded by the polynucleotide, a vaccine composition comprising said polypeptide and a polynucleotide having a sequence SEQ ID NO 1.

2. Claims: 1-10,15,17-23,26,29,32-48 (all partially)

Same as invention 1, but directed at polypeptides of clone 2 (SEQ ID 7) with sequences SEQ ID NO 8-12.

3. Claims: 1-10,15,17-23,26,29,32-48 (all partially)

Same as invention 1, but directed at polypeptides of clone 3 (SEQ ID 13) with sequences SEQ ID NO 14-21.

4. Claims: 1-10,15,17-23,26,29,32-48 (all partially)

Same as invention 1, but directed at polypeptides of clone 4 (SEQ ID 22) with sequences SEQ ID NO 23-26.

5. Claims: 1-10,15,17-23,26,29,32-48 (all partially)

Same as invention 1, but directed at polypeptides of clone 5 (SEQ ID 27) with sequences SEQ ID NO 28-31.

6. Claims: 1-10,15,17-23,26,29,32-48 (all partially)

Same as invention 1, but directed at polypeptides of clone 6 (SEQ ID 32) with sequences SEQ ID NO 33-36.

7. Claims: 11-14,16,24,25,27,28,30,31 (all completely), 1-10, 15,17-23,26,29,32-48 (all partially)

Same as invention 1, but directed at polypeptides of clone 7 (SEQ ID 37) with sequences SEQ ID NO 38-41.

INTERNATIONAL SEARCH REPORT

information on patent family members

International Application No

PCT/CA 99/00114

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9818931 A	07-05-1998	AU 5194598 A	22-05-1998
		AU 6909098 A	22-05-1998
		EP 0942983 A	22-09-1999
		EP 0941335 A	15-09-1999
		WO 9818930 A	07-05-1998
